Aviation Human Factors Division
Institute of Aviation

University of Illinois
at Urbana-Champaign
1 Airport Road
Savoy, Illinois 61874

Workload Demands of Remotely Piloted Vehicle Supervision and Control: (I) Single Vehicle Performance

Christopher D. Wickens
and Stephen Dixon

Technical Report AHFD-02-10/MAD-02-1

September 2002

Prepared for

Micro Analysis and Design
Boulder CO

Contract ARMY MAD 6021.000-01

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **SEP 2002**	2. REPORT TYPE	3. DATES COVERED **00-00-2002 to 00-00-2002**

4. TITLE AND SUBTITLE **Workload Demands of Remotely Piloted Vehicle Supervision and Control: (I) Single Vehicle Performance**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Illinois at Urbana-Champaign,Aviation Human Factors Division,Savoy,IL,61874**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Eighteen licensed pilots flew a remotely piloted vehicle (RPV) simulation on three 10-leg missions. The simulation required navigating to an enemy target, monitoring for unexpected targets in a 3D image display, and monitoring on-board systems. Both of the first two tasks required zooming in to inspect 3D images of the targets. Displays to support these activities were presented on a 19 inch Screen. Each pilot flew in a baseline condition, a condition supported by redistributing some information to the auditory modality, and a condition supported by automating much of the navigational tasks. The results revealed considerable interference between the tasks components. Some aspects of this interference were relieved by auditory offloading. However other aspects were not, suggesting that heavy cognitive demands of image manipulation dominated any benefit for using separate perceptual modalities. Navigational automation also relieved some aspects of task interference. The results are interpreted in terms of their relevance to different theories of multiple task interference, and to the concept of "cognitive tunneling".

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**39**	

Abstract

Eighteen licensed pilots flew a remotely piloted vehicle (RPV) simulation on three 10-leg missions. The simulation required navigating to an enemy target, monitoring for unexpected targets in a 3D image display, and monitoring on-board systems. Both of the first two tasks required zooming in to inspect 3D images of the targets. Displays to support these activities were presented on a 19 inch Screen. Each pilot flew in a baseline condition, a condition supported by redistributing some information to the auditory modality, and a condition supported by automating much of the navigational tasks. The results revealed considerable interference between the tasks components. Some aspects of this interference were relieved by auditory offloading. However other aspects were not, suggesting that heavy cognitive demands of image manipulation dominated any benefit for using separate perceptual modalities. Navigational automation also relieved some aspects of task interference. The results are interpreted in terms of their relevance to different theories of multiple task interference, and to the concept of "cognitive tunneling".

Introduction

The Army is currently interested in increasing the number of remotely piloted vehicles (RPVs) that are available to do surveillance, as these RPVs (Army Hunter and Shadow), and others have seen valuable service in Bosnia and Afghanistan. The challenge is that there are a limited number of RPV pilots, and, currently, more RPVs cannot be flown, without more trained pilots. This is a particular manpower challenge because each RPV is currently flown with two pilots. Thus the Army would like to progress from the current situation of 1 RPV / 2 pilots to one involving 2 (or possibly more) RPVs / 1 pilot.

On the one hand, it is easy to see how this change could drastically increase the workload, perhaps to a point where missions are sacrificed. On the other hand, appropriate harnessing of automation and interface design could prevent such a downfall. Indeed, in aviation, the Boeing Company successfully convinced the FAA in the 1980's that it was possible to downsize the crew complement on large jet transports, from 3 to 2, without increasing the workload, by the appropriate implementation of automation.

The long term objective of research carried out in our laboratory, as part of the Human Machine Interface workload measurement component, is to model and assess the workload imposed on a single pilot controlling multiple RPVs. In support of this objective, it was necessary first to develop an RPV simulation, and evaluate pilot performance using such a simulation in single RPV conditions. The current report describes this effort. As the current configuration of operator control of RPVs suggests, even control of a single vehicle imposes multiple concurrent tasks (in current operations distributed between two operators). Thus the current research examines the interference between these tasks within a single RPV for a single operator.

In addition the current report examines two techniques designed to reduce some of the time sharing requirements of single RPV control: via offloading many visual demands to auditory channels, and via automating certain components of the task. Thus a second purpose of the current research is to examine the effectiveness of these two techniques. To the extent that they are effective in reducing resource overload in single RPV control, they should be more so in the control of multiple RPVs.

A third purpose of the current experiment is to provide data that could be used to help validate computational workload prediction models such as IMPRINT, that could then be used to predict the workload implications of other interface alterations. Different versions of such models of workload overload (performance breakdowns) in multi-task environments are based on three fundamentally different mechanisms: scheduling, single resources (task demand) and multiple-resources (Sarno & Wickens, 1995; Wickens & Hollands 2000). In the following, we briefly review both the theoretical and applied work that addresses these issues, before describing closely related research to the multiple-RPV experiment carried out here.

Theories of Multiple Task Performance

Theoretical positions on how humans perform two tasks with concurrent demands, differ in the extent to which they allow activity for the two tasks to actually be carried out in parallel, rather than in sequence. Models of the latter class are often described as **single channel** models of task performance (Hendy, Liao, & Milgram, 1997; Liao & Moray, 1993; Welford, 1967; Pashler, 1998). They assume that some aspect of performance or information processing on one task cannot be carried out in parallel with same or different aspects of processing on a second task, so that these aspects must be scheduled in sequence. In the extreme, single channel theory posits that one entire task must be completed before another can be started. Some versions that are less extreme allow switching between partially completed tasks. Other versions posit that only certain aspects of information processing are single channel, such as access to foveal vision (Liao & Moray, 1993; Moray, 1986), the selection of responses (Pashler, 1998), or the utterance of speech responses (Wickens, 2002). Sometimes these single channel stages of processing are referred to as "bottlenecks" (Keele, 1973).

Single channel models thus need to account for two general aspects of performance: **what** aspect of performance and processing defines the single channel (e.g., foveal vision, response selection, vocalization, the entire task, etc.), and when there is conflict, **how** some executive processor or scheduler manages the conflict: which task is performed first? how frequently, and in what conditions can the second-performed task interrupt the first? how frequently is switching between the tasks carried out? As such, this aspect of sequential task performance is often characterized by the term "task management" (Pattipati & Kleinman, 1991; Raby & Wickens, 1994; Meyer & Kieras, 1997; Chou, Madhavan, & Funk, 1996; see Wickens & Hollands, 2000 for a review of issues). Since foveal vision is often considered a restricted bottleneck, and the allocation of that vision to tasks is easily measured via scanning, visual scanning models have often been applied to understanding task scheduling in supervisory control environments, such as that encountered in the control of vehicles (Carbonnell, Ward, & Senders, 1968; Moray, 1986; Wickens, Helleberg, Goh, Xu, & Horrey, 2001).

In contrast to single channel models, **parallel processing models** address the circumstances in which sequential scheduling need not be carried out, because both tasks are sufficiently low in their demand for limited resources that they can be accomplished in parallel (even though such parallel processing may not be perfect). Kahneman (1973), Moray, (1967) and Norman and Bobrow (1975) were early advocates of such a view that the human had a limited supply of "attentional resources" that could be allocated, in graded quantity between two concurrently performed tasks. Tasks that were of higher priority, or of greater difficulty received a greater proportion of this resource allocation relative to tasks of lower priority or of lesser difficulty. In the extreme, a task that requires minimal resources for its performance is said to be "automated" (Schneider & Shiffrin, 1977). Such models nicely account for the tradeoffs between, for example driving and conversing, or flying and radio operations. In both cases, concurrent performance is clearly feasible. As a consequence, the driver or pilot does not need to "schedule" times in which vehicle control is done in the absence of verbal activity, or times in which verbal activity is done and vehicle control ceases. Instead, the operator can clearly be talking while driving or flying. The operator can however place relatively more emphasis on one activity or the other (priority-based resource allocation). Furthermore, if driving becomes quite difficult, the quality conversation declines, as more resources are shifted to driving, away from conversing. If the conversation becomes very taxing or engaging, the quality of driving may suffer as more resources are shifted to conversing.

Parallel processing resource models such as those proposed by Kahneman (1973) and Norman and Bobrow (1975) did not distinguish between different kinds of resources, and hence were labeled **single resource models**. Subsequent research revealed that there were substantial differences in task interference (workload overload costs) associated with qualitative changes in task structure, in addition to task allocation priorities or quantitative demand (Navon & Gopher, 1979; Wickens, 1980). These results led researchers to posit **multiple resources** within the human processing system. In the above driving example, concurrent processing might be quite effective while the operator is listening to conversation (auditory input), but if he must read a printed transcript of the same conversation, concurrent performance will be degraded, even if it is not entirely abandoned (switching to a single channel mode). Such a finding, observed in a wide range of research (see Wickens, 1980; Wickens & Hollands, 2000; Wickens, Goh, Helleberg, & Talleur, 2002), allows the assumption that the auditory and visual channels of perception define separate resources. In a highly visual environment, altering the display of one task from vision to audition should reduce the amount of task interference. While there are many other nuances in multiple resource theory, and other dimensions defining multiple resources, these will not be described here (see Wickens, 2002).

It is important to note that single channel theory, single resource theory and multiple resource theory are not mutually exclusive. Rather, multiple task performance in complex environments can often be described in terms of the conditions that allow one theoretical mechanism or the other to better account for certain aspects of task performance. Thus for example, when concurrent performance **is** possible, conditions may differ from each other primarily in terms of the demands (e.g., the level of turbulence through which a vehicle is traveling). In this case, single resource models are adequate. In other circumstances, comparisons may need to be made between different task interfaces that impose greater or less competition for multiple resources; for example as auditory versus visual displays are compared (Wickens et al., 2002), or voice versus manual responses are compared (Sarno & Wickens, 1995). In such cases multiple resource models are appropriate. In still other cases, because of either excessively high resource demand (single resource theory), or a change in task structure that prohibits concurrent processing (using one, rather than two resources), the operator may "regress" to a single channel form of behavior, in which task scheduling is the prominent mechanism accounting for multiple task performance, and single channel model assumptions about scheduling become most relevant.

One particularly relevant aspect of such regression (to single channel performance) is a form of behavior often observed under high multitask workload, that may be described as "cognitive tunneling" or "task fixation". These are circumstances that (a) lead the operator to focus attention only on one task (single channel behavior), and (b) leave concurrent tasks neglected or unattended for a non-optimally long period of time; that is, the operator does not switch attention back to the unattended task, as frequently as he or she should, given its priority or importance (Moray & Rotenberg, 1989; Fadden, Ververs, & Wickens, 2001; Kerstholtz, Passenier, Houttuin, & Schuffel, 1996).

Models of Workload and Time Sharing

Various researchers have proposed computational models to account for multiple task performance, based on differing combinations of the three theoretical mechanisms described above . A series of “queuing models” have been developed to account for single channel behavior in dynamic multi-task or multi-channel environments (Hendy et al., 1997; Moray, 1986; Pattipati & Kleinman, 1991). Meyer and Kieras (1997) have developed decision rules for task management multi-task environments in relatively basic laboratory tasks. Few computational models have been developed from single resource theory, perhaps because of the difficulty in quantitatively defining “resource demand”. However some of these resource demand assumptions have been incorporated into a series of computational versions of multiple resource models. Many of these are based, initially upon the concepts of a WINDEX model developed by North and Riley (1989; see also Laughery, 1989 for a similar concept). Such models penalize concurrent task performance jointly for excessive resource demand, and for demand on overlapping resources. Both scheduling (single channel assumptions) and multiple resource concepts have been incorporated into two recent modeling efforts, MIDAS, developed at NASA Ames Research Center, and Windcrew/IMPRINT, developed jointly by the Army Human Engineering Lab, and MicroAnalaysis and Design (Laughery & Corker, 1997).

Despite the potential importance of such models in predicting human performance breakdowns in overload situations, few studies have carried out competitive evaluations of different models on the same data set (or competitive evaluations of models, making different assumptions about time sharing). Two such approaches (Sarno & Wickens, 1995; Liao & Moray, 1993) have both provided different forms of evidence on the relative value of different model assumptions. Sarno and Wickens, using relatively simple tasks that mimicked the demands of flying and concurrent task management, found that pure single channel models provided a poor fit for the data, and that multiple resource assumptions were quite important. They also observed that single resource assumptions (i.e., modeling task difficulty or resource demand) were relatively less critical, as long as it was assumed that a more demanding task required longer to complete. Liao and Moray used a two and four task simulation that had many features in common with the RPV task to be examined here, and observed that single channel theory did a reasonably good job of accounting for four task behavior (because of very restricted limitations of foveal vision), but that when subjects only shared two tasks, additional factors, not accounted for by scheduling, became relevant. These factors related to either single or multiple resources, although the investigators did not discriminate between them.

The intention then of the experiment we report below is to provide data on how well pilots fly an RPV simulation in an unaided form, and in a form involving both automation and auditory offloading of some of its components. Such data can be used to assess the circumstances under which single channel behavior and multiple-resource mechanisms are manifest, providing validation data for subsequent applications of computational workload modeling. In the simulation, based upon a cognitive task analysis of Hunter/Shadow pilots, carried out at Ft Huaccucca, Ariz., pilots performed three general tasks: (1) navigating their RPV to a series of “command targets” at which reports of enemy activity (based on an image camera) were required; (2) monitoring the terrain over which they flew for any “targets of opportunity” that might appear, and (3) monitoring on board system status indicators for various indications of abnormalities, described as “system failures”. As such, the simulation required fairly extensive time sharing between the different phases of each of these three tasks, as detailed below. The simulation was flown in a baseline condition, and conditions associated with auditory offload and navigational automation.

Methods

Participants

Participants were 16 male and 2 female undergraduate students (ages 18-25) enrolled in the University of Illinois Aviation Program. All the participants had at least a visual flight rules (VFR) rating, with some instrument flight rules (IFR) experience. All participants received $8 per hour for their time.

Apparatus

Participants were asked to perform three simultaneous tasks: 1) a tracking/navigation task, 2) a target-search task, and 3) a monitoring task for system failures during three 50-70 minute missions. The tasks were displayed on a Hitachi CM721F 19-inch monitor (37-degrees visual angle) with 1280x1024 resolution, and full-scene anti-aliasing and 32MB texture memory. The experiment was run using an Evans & Sutherland SimFusion 4000q dual 1Ghz PIII processor with an OPENsim Graphics card.

As seen in Figure M1, the experimental environment was subdivided into four separate windows. Figure M2 contains the range of visual angles between the individual windows.

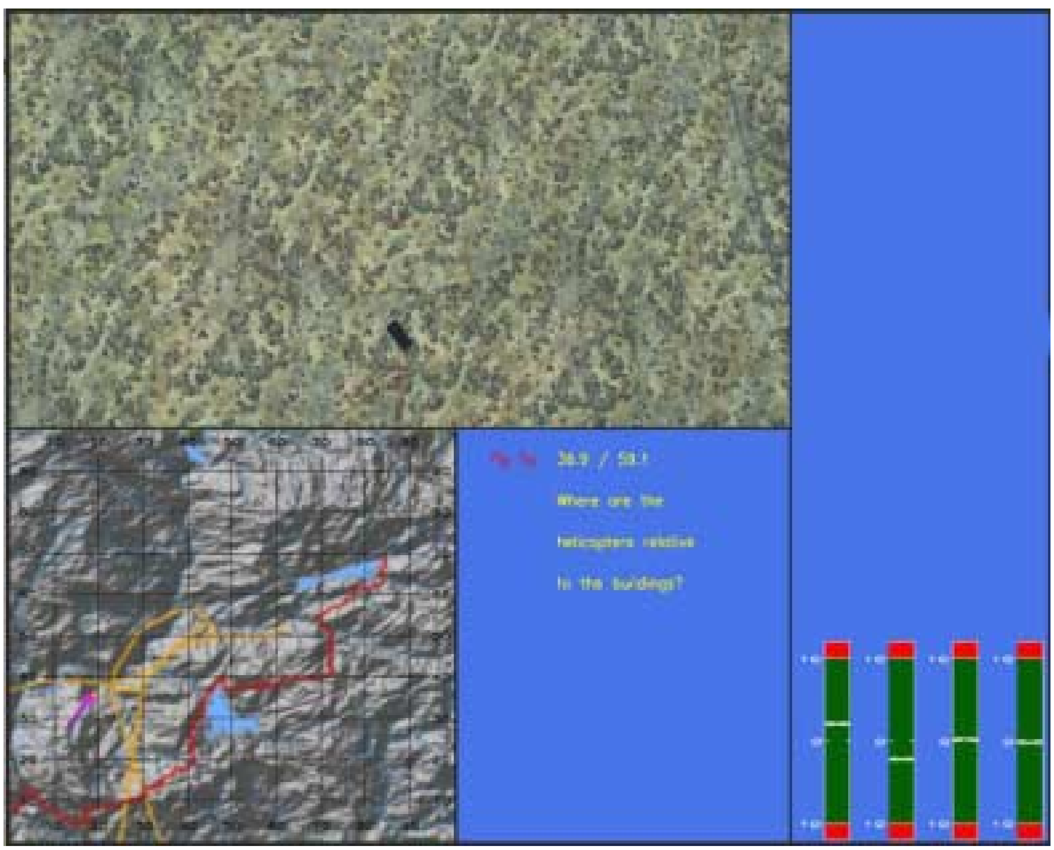

Figure M1. A screenshot example of the experimental display. Actual display was larger and more legible than this figure rendering.

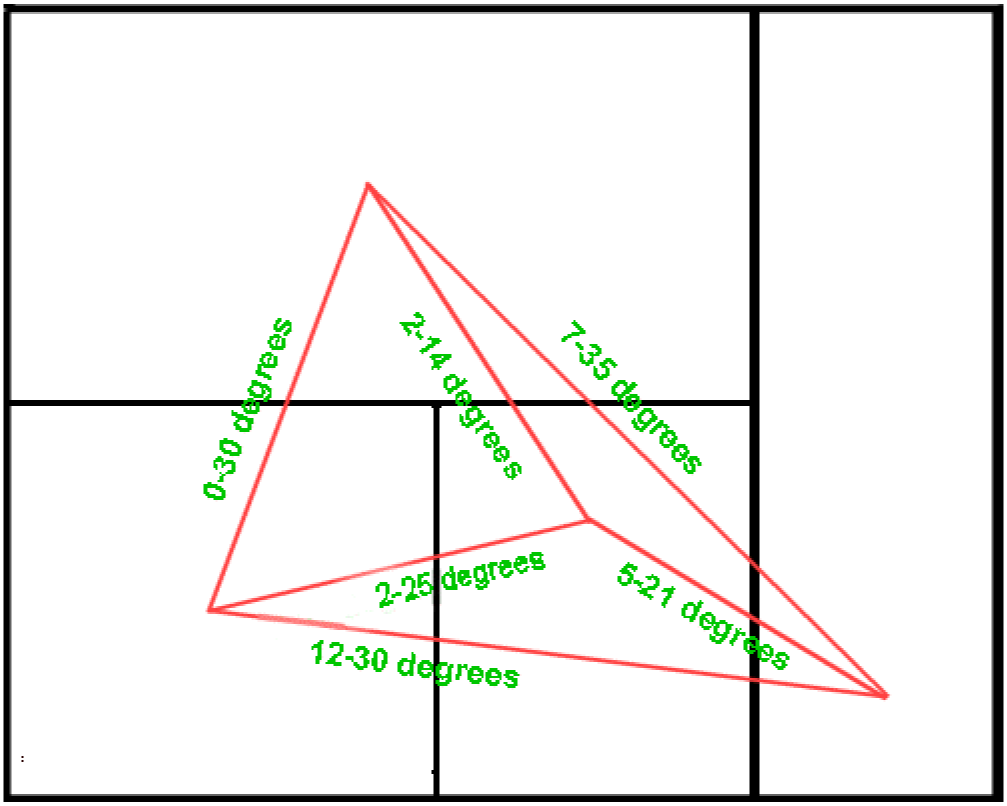

Figure M2. Range of visual angles between the four main windows / tasks. The ranges go from the two farthest points of interest to the two closest. The average visual angle is equal to the visual angle between the center points in the diagram.

The top left window contained a 3D egocentric image view of the terrain forward and/or below the RPV. The sample figure shows a command target (CT) at normal viewing distance (i.e., 6000 feet altitude). The ability to manipulate this view depended on whether the operator was tracking a straight line path or loitering around a target. During regular tracking periods, the operator could view straight down to the ground or pan the camera up 45 degrees towards the horizon by manipulating the camera's y-axis. The momentary camera angle was indicated by a yellow dot along a red y-axis bar at the bottom center of the 3D window. During a loiter pattern, the operator was able to extend the viewing angle from 0 to 90 degrees along both the x- and y-axes. A zoom feature (up to 100x) was also available only in the loiter pattern.

The bottom left window contained a 2D top-down map of the 20x20 miles simulation world. Coordinates (which formed a grid) from 0-100 were placed along the x- and y-axes for navigation purposes. The yellow and red lines denoted minor and major roads, respectively. The smaller blue lines denoted rivers, and the large blue shapes denoted lakes.

The bottom center window contained the Message Box, with "fly to" coordinates and CT report questions. These instructions were present for 15 seconds at a time. During the

Automation condition, the lower half of this box also contained a place to type in the coordinates for the next CT.

The bottom right window contained the four system failure (SF) gauges. Each gauge represented a different onboard system. The white bars oscillated up and down continuously, each driven by sine waves ranging in bandwidth from 0.01 Hz to 0.025 Hz.

A SF occurred when one of the white bars moved gradually into a red zone.

Participants used a Logitech Digital 3D joystick to manipulate the aircraft/camera and a X-Key 20-button keypad with which to indicate responses. As seen in Figure M3, the joystick had controls for turning the UAV, manipulating the camera on the x- and y-axes, zooming, detecting targets, loitering around targets (to the left or right), and detecting SFs. The keypad was used for indicating which system failure occurred, the ownship coordinates for that system failure, and for typing in mission coordinates during the Automation condition. The experimenter used a separate keypad to record correct or incorrect responses and to indicate when the operator detected a target of opportunity (TOO) or a command target (CT).

Figure M3. Logitech joystick and X-Keys keypad.

Conditions and Tracking Task

There were three conditions: 1) Baseline, 2) Auditory, and 3) Automation. In each condition, the participants were seated approximately ½ meter away from the screen and required to fly one mission, which consisted of 10 legs with command targets at the end of each leg. Figure M4 shows a sample mission, highlighting the command targets and flight paths of the 10 legs. Each leg was approximately 10-12 kilometers long, and took approximately 5-8 minutes to complete. Each mission used a different set of legs and targets, so there were no repeating maps or targets for a participant.

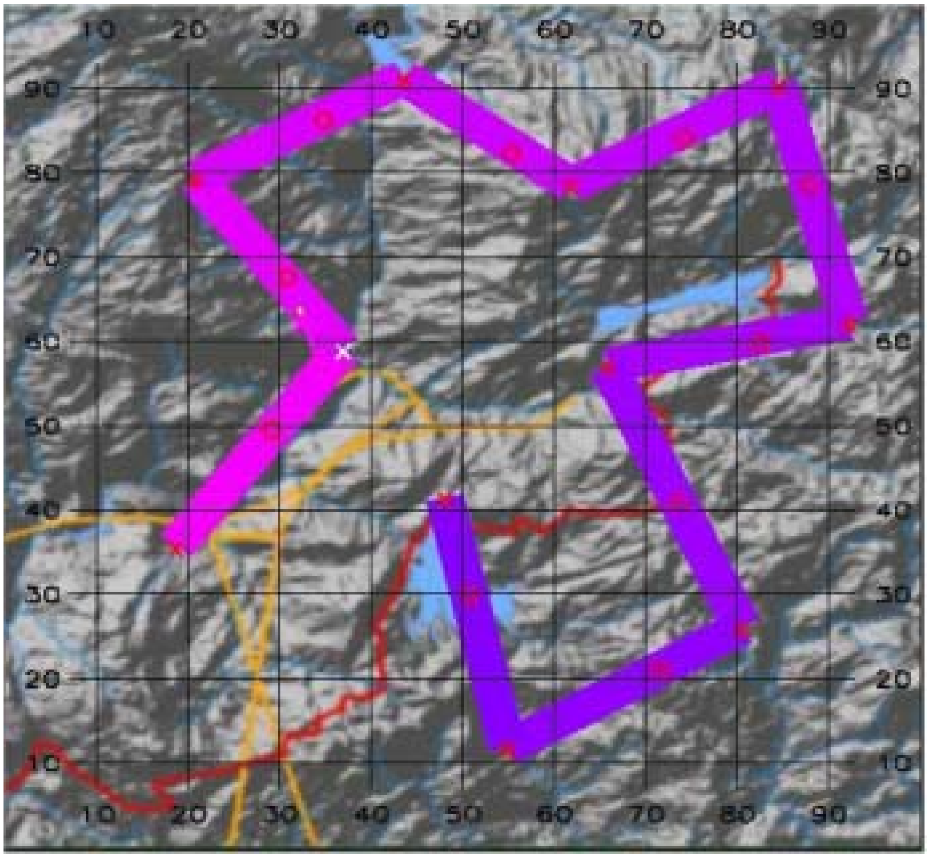

Figure M4. A sample mission. CTs are designated with red X's while TOOs are designated with red O's. This preplanned route was not shown to participants. The wide path along each leg indicates the width of the image in the 3D view when at minimum zoom.

There were two different forms of tracking control: 1) manual mode, and 2) automatic-mode. During the Baseline and Auditory conditions, the participants were required to manually control the UAV heading through each mission. This first-order control was accomplished by twisting the joystick to the left or the right. There was no disturbance in the control; that is, if left alone, the UAV would travel in the straight line established by the twist without deviation from its path. Participants were not responsible for airspeed (fixed at 70 knots), or altitude (fixed at 6000 feet). The operator did not have the capability to pitch, bank, or roll the UAV.

During the Automation condition, the operator was *not* responsible for manually tracking the UAV. Instead, he or she was required to type in the mission coordinates of the next command target at the beginning of each leg, using the keypad. The computer then automatically guided the UAV along a direct, straight-line path to those coordinates.

Both the Baseline and Automation conditions entailed visually reading all instructions and system parameters. The Auditory condition presented auditory instructions and alarms for system failures (see below).

Target Searching and Reporting Tasks

A command target (CT) was located at the end of each mission leg, at the coordinates specified at the leg beginning. As seen in Figures M1 and M5, which depict a typical CT at 0x zoom and again at 100x zoom, respectively, these were very salient and easy to find. They consisted of a building (e.g., warehouse, factory, hanger, etc.) with 1-3 tanks and/or helicopters located within 10-50 feet around them. These weapons were always located on the north, south, east, or west sides.

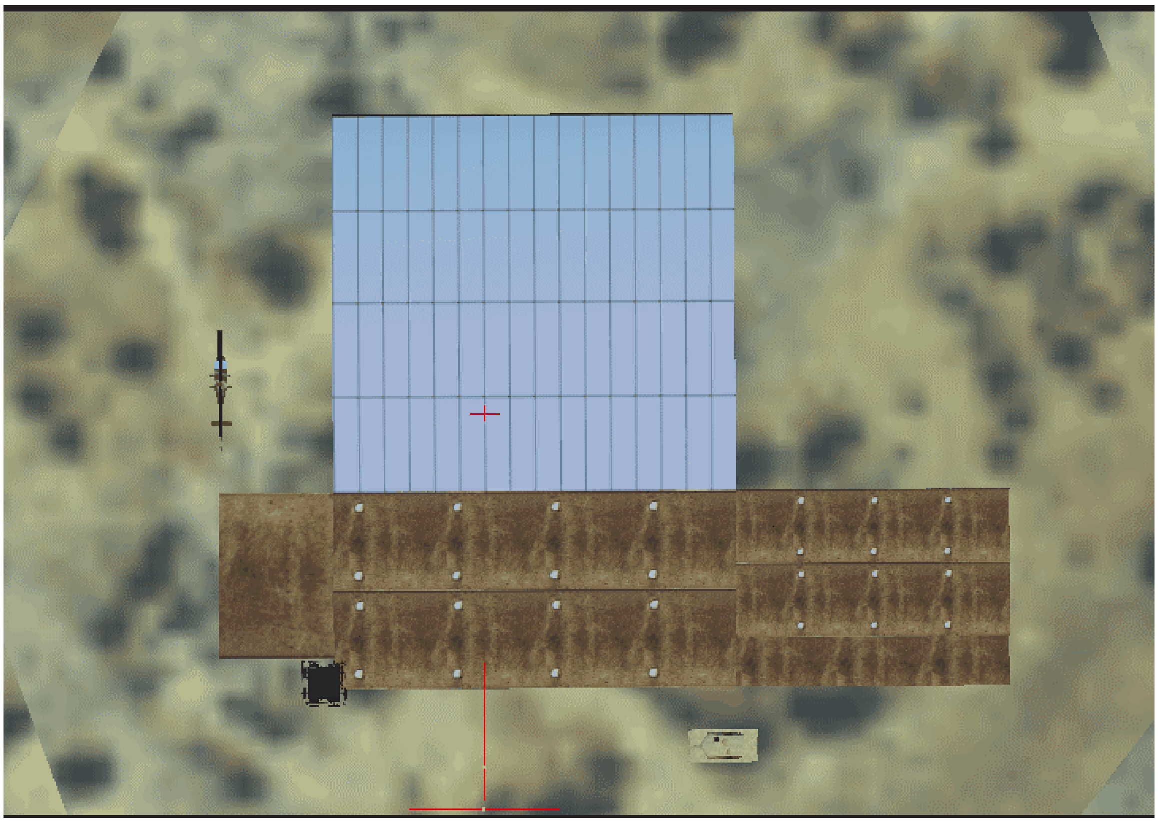

Figure M5. An example of a command target under a 100x zoom, from an angle looking directly downward (-90^{o}).

The pilots were required to loiter around all CTs, zoom in the camera for a closer view, and respond to questions that appeared in the message box (or were spoken in the Auditory condition) about what they could see. Sample questions might be: 1) How many tanks are there and where are they located in relation to the building?, 2) Report the number of weapons present, or 3) Where are the helicopters located? The questions were divided into two main categories: a) questions requiring cardinal direction judgments, and b) questions requiring counting the number of a type of weapon. These questions could only be answered once the operator had zoomed in close to the CT.

These questions were offered once at the beginning of each leg and stayed visible in the message box for 15 seconds (in the manual and automatic modes) or were presented aurally by digitized speech in the auditory mode. If the pilots forgot the question, they were allowed to hit a Repeat button on their keyboard at any time. The number of repeats were recorded.

Along each leg, pilots were also instructed to search for Targets of Opportunity (TOO). Figures M6 and M7 depict a typical TOO at 0x zoom and again at 100x zoom, respectively. As seen in Figure M6, these TOOs were camouflaged and difficult to see at 6000 feet. They occupied between 1-2 degrees of visual angle, and could not generally be detected unless foveated. All TOOs were the same basic square "bunker" shape and came in three sizes. There was one TOO per mission leg, which was located randomly somewhere in the middle 60% of each leg (i.e., between 20% and 80% of distance traveled); however, participants were not told this. They were only told that the TOO was somewhere along the direct-line path between CTs. Around each TOO were 1-3 tanks and/or helicopters, located within 10-30 feet of the bunker. These weapons were always located on the north, south, east, or west sides.

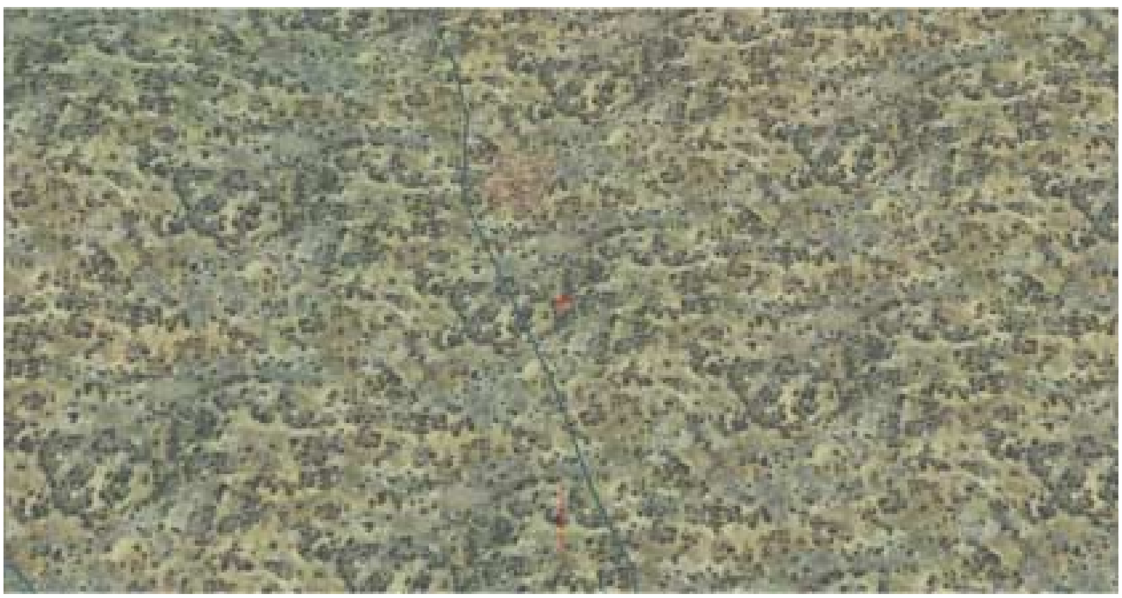

Figure M6. An example of a medium-sized camouflaged TOO at 0x zoom. The TOO is located just to the left of center, about 30% of the way down from the top. The target was actually somewhat more visible in the dynamic color screens than in the current rendering.

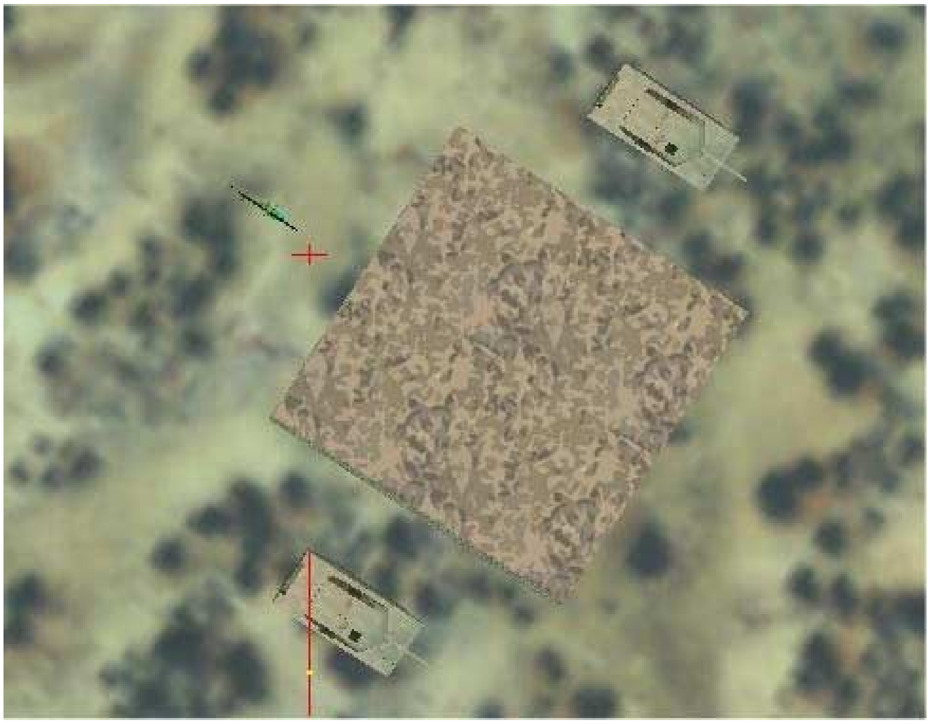

Figure M7. An example of a Target of Opportunity (TOO) under a 100x zoom.

The question for TOOs was always the same: “what weapons do you see and where are they located?” As with the CTs, these questions could only be answered once the operator had zoomed in close to the target.

If the participant detected a CT or TOO, he or she was required to indicate detection by pulling the joystick trigger. The experimenter then pressed the appropriate button on the experimenter’s keypad to indicate whether this was a TOO or CT. After deciding that the UAV was close enough to the target to begin inspection, the participant pressed the loiter button (loiter would be selected either left or right) on the joystick (see Figure M3). This put the UAV into an automated “racetrack” pattern around the target. This racetrack pattern was 1.3 kilometers wide and 2.1 kilometers long, and took between 2.5 to 3 minutes to complete an entire 4.8-kilometer circuit. The UAV turned 3 degrees per second at the ends of the oval. During the loiter pattern, the participant was able to use the x- and y-axes of the camera, as well as to zoom in and focus more closely on the target. The task of keeping the TOO in view while zooming *and* moving *and* keeping track of cardinal directions, was extremely challenging.

After making the report, the participant could then hit the loiter button again, which would unloiter the UAV and unzoom the camera, returning the egocentric view to 6000 feet altitude. In the Baseline and Auditory conditions, once the report was completed, the participant had to relocate CT coordinates and reorient the UAV to the direct path to the CT. In the Automation condition, the UAV automatically resumed the correct path to the next CT. The duration of time between detection and completion of the final report was recorded, as the response time measure. There was no separate measure of detection time (i.e., the time between the appearance of the TOO on the screen, and the pilot's depression of the "detect" button, since the former event was difficult to establish on a case by case basis, given variability of the course, and the camera angle).

System Monitoring Task

During each mission, participants were also asked to detect system failures (SF). A SF occurred when a system gauge needle went out of bounds (i.e., passed from the green zone into the red zone at either the top or bottom of the gauge; see Figure M1). Each SF lasted approximately 30 seconds before automatically resetting (i.e., moving back into the green zone) if not detected. Each mission consisted of 8 SFs (2 per gauge); that is, not every leg contained a SF, and no leg contained more than one SF, although participants were not told this. The number of correct detections and the time it took to detect the SFs were recorded.

If a SF was detected, the participant pressed a "detect" button on the joystick. Then he or she pressed the appropriate button on the keypad to indicate which system had failed. Lastly, he or she typed in his or her current ownship coordinates and then hit Enter. The duration of this time between detection and final report completion was recorded.

The SFs were categorized under 5 types: A) during initiation of flight heading (i.e., while the pilot was consulting the message box and the 2D map, deciding how to turn the plane, and establishing the correct course); B) during regular flight, when no TOO was visible; C) approximately 5 seconds after a TOO entered the field of view; D) approximately 5 seconds after a TOO loiter pattern was entered; E) approximately 5 seconds after a CT loiter was entered. These will be referred to below as SF_A, SF_B, SF_C, SF_D, SF_E. Figure M8 shows a typical mission leg and where SFs might occur along the leg. As noted above, a maximum of one SF occurred per leg.

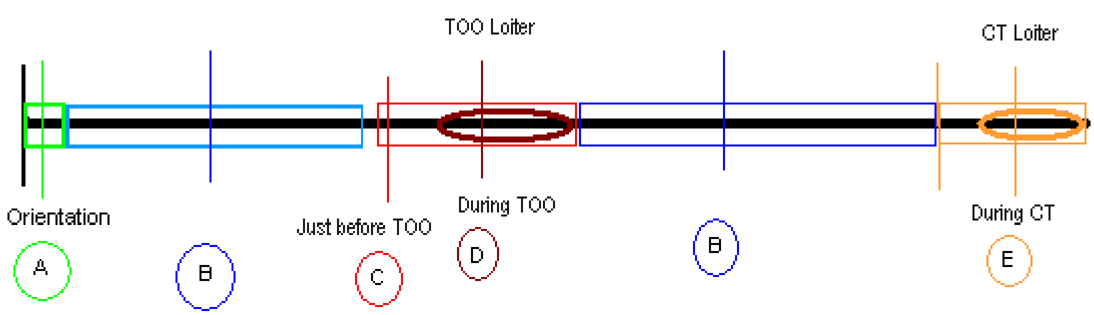

Figure M8. A timeline of SFs for a typical mission leg.

Procedure

Each participant was seated in a comfortable chair in front of the mission monitor. After signing the consent form, participants were asked to read the instructions for the experiment. Once they completed the instructions, they were allowed to spend 10-12 minutes on a practice mission, during which they would be exposed to two CTs, two SFs, and one TOO. Any questions they might have were answered verbally by the experimenter. Once the practice mission was completed, the participant was asked if he or she felt comfortable with the controls and instructions. All the participants responded positively and none of them asked for more time.

The experimenter then started the first mission. Immediately, the instructions (i.e., "fly to" coordinates and CT question) for the first mission leg were posted visually (for Baseline and Automation conditions) in the Message Box, or aurally (for the Auditory condition). The participant then either manually (in the Baseline and Auditory conditions) oriented the UAV along the flight path towards the coordinates of the next CT, or typed in the coordinates (in the Automation condition) using the keypad.

As mentioned previously, if a SF occurred, the participant indicated this by pressing the appropriate button on the joystick, followed by pressing the appropriate button on the keypad to indicate which SF had failed, followed by typing in the current ownship coordinates and then hitting Enter. These three steps were required for all SFs, and the detection rates and response times were recorded.

Upon detecting a TOO, the participant was required to loiter around it and zoom in the camera for a closer view, and then report all weapons present and where they were located. Detection was indicated by pulling the joystick trigger. In order to loiter, the participant hit the loiter button on the joystick. This put the UAV into the automated racetrack pattern around the target described above. The participant then was able to zoom in and focus more closely on the target.

As mentioned earlier, the participant was required to describe verbally which weapons were present and their cardinal locations around the building. Once the report was made, the experimenter hit the "correct" or "incorrect" button on the experimenter's keypad (note that these buttons were not labeled so the participant would not be aware of his or her accuracy). The participant could then unloiter and continue on the flight path towards the next CT.

If a SF was detected during a TOO, the participant was free to choose his or her own order of importance. The two procedures were programmed to allow timesharing and overlapping of actions. That is, for example, the participant could begin a loiter, then hit the "detect" button for a SF, then zoom in and report on the TOO, and then finish with the SF, or any other combination of actions. He or she was free to do these in a sequential order or to do some of the tasks simultaneously.

Upon detecting a CT at the end of each leg, the participant followed the same basic procedure as with TOOs, with two differences: 1) the questions for the CTs were always different, and 2) when the participant hit the "unloiter" button at the end of the report, the instructions for the next mission leg and CT would appear.

This process was repeated throughout all 10 legs of each mission.

Design

Using a within-subjects design, all 18 participants were exposed to each of the three conditions. The conditions and missions/maps were counterbalanced. Since there were three different maps (A, B, & C), defined by their legs, target locations, and target types, these were crossed with display condition. Thus 6 participants experienced Map A in the Baseline condition, 6 different participants experienced Map B in the Baseline condition, and 6 different participants experienced Map C in the Baseline condition; and similarly for the Auditory and Automation conditions.

Dependent variables included: 1) tracking error; 2) detection rates, response times, and accuracy for SFs and TOOs; 3) response times and accuracy for CTs; and 4) repeats.

To review, the three major tasks are shown in Table M1. Each task is broken down into a series of subtasks, that typically appear in sequence.

Table M1. Task analysis.

1. Navigation:
   - 1.1 Read (or hear) CT location
   - 1.2 Establish coordinates (by orienting vehicle by joystick control or typing)
   - 1.3 Monitor heading toward CT location on 2D map (and re-orient if necessary)
   - 1.4 Refresh memory for location and final report
   - 1.5 Inspect image
     - 1.5.1 Enter loiter
     - 1.5.2 Zoom in
     - 1.5.3 Adjust camera orientation
     - 1.5.4 Count identify and/or assess cardinal orientations
     - 1.5.5 Verbal report of content.

2. TOO task:
   - 2.1 Monitor 3D display
   - 2.2 Inspect image if target located (see 1.5 for subtasks)

3. System Failure
   - 3.1 Monitor for System failures
   - 3.2 Identify failure
   - 3.3 Keyboard data entry

Results

This results section has been divided into four main subsections: 1) Tracking error, 2) System Failures, 3) Target Searching (command targets and targets of opportunity), and 4) Repeats. Although these subsections are interrelated in many ways, the statistical results will be analyzed separately. The relationships between these sections, as well as the models which predict the results found here, will be explained further in the discussion section.

Unless specified otherwise, all data were analyzed using the within-subjects design. Occasionally, missing data points (e.g., if a participant missed a target of opportunity, they would also miss the system failure (SF) that accompanied those targets) required the following statistical approach in order to preserve the remaining data: if the target (TOO or SF) was possible to detect, but was not detected, then the maximum value for RT was allotted to that subject for that trial. For example, subjects had 30 seconds maximum to detect a SF. If they failed to detect the SF, then the RT was set at 30 seconds.

Because our interests lay in determining the effects that auditory and automation offloading have on the operator's performance, our statistical comparisons focused mostly on differences between the Baseline and Auditory conditions, and differences between Baseline and Automation conditions. We were not particularly interested in analyzing differences between the Auditory and Automation conditions, so the all-encompassing ANOVA was generally eschewed for the two specific comparisons. All the contrasts in this experiment were planned a priori and orthogonally. For similar reasons, we also made no adjustments (e.g., Bonferroni) to control familywise Type 1 error rates (see Keppel, 1982, for more discussion on this approach).

Tracking Task

Mean absolute error (MAE). MAE was only analyzed between the Baseline and Auditory conditions, since the Automation condition purposely had no tracking error (i.e., in this paradigm, the auto-tracker perfectly followed the straight line trajectory between command targets and the MAE was zero). A paired t-test, [$t(17) = .47$, $p = .32$], showed no significant difference between the overall Baseline and Auditory tracking means.

Further analysis between the first five legs and the last five legs of each mission showed no practice effects; that is, there was no improvement in tracking performance over time. In fact, there was a significant *decrement* in tracking performance from the first five legs to the last five legs in the Baseline condition, [$t(17) = 5.80$, $p < .0001$], or a 61% increase in tracking error, and in the Auditory condition, [$t(17) = 4.84$, $p < .0001$], or a 55% increase in tracking error. This may be caused by one (or more) of three influences: 1) fatigue, 2) loss of concentration, or 3) a dual-task tradeoff that changes over the course of the experiment, to favor monitoring at the expense of control.

System Monitoring Task

System failures (SF) detection rate. As described in the Methods section, the system failures were categorized under 5 types: A) during initiation of flight heading (i.e., while the subject is deciding how to turn the plane, and establishing the correct course); B) during regular flight, when no TOO is visible; C) just after a TOO enters the field of view. We do this to establish if dealing with the SF will disrupt monitoring; D) just after a TOO loiter pattern is entered. We do this to establish if loitering/image inspection will disrupt system monitoring; E) just after a command target loiter is entered. Figure R1 shows a timeline of a typical mission leg and where these SFs might be located.

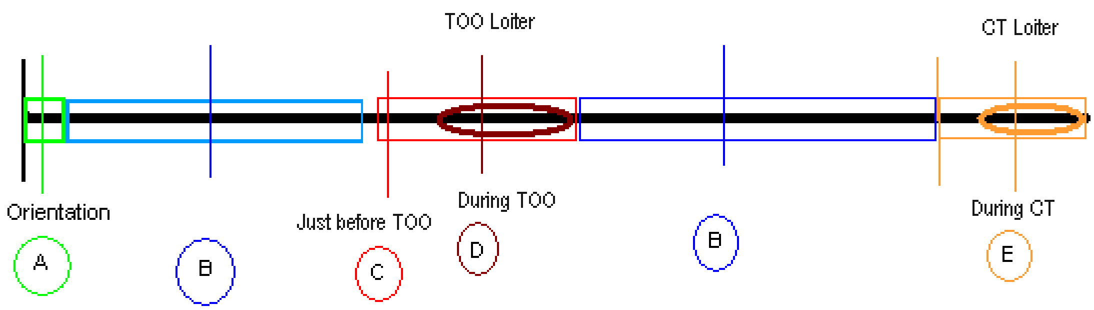

Figure R1. A timeline of SFs for a typical mission leg.

Baseline vs. auditory. The data in Figure R2 presents the participants' detection rate for SF_A through SF_E, across the Baseline and Auditory conditions.

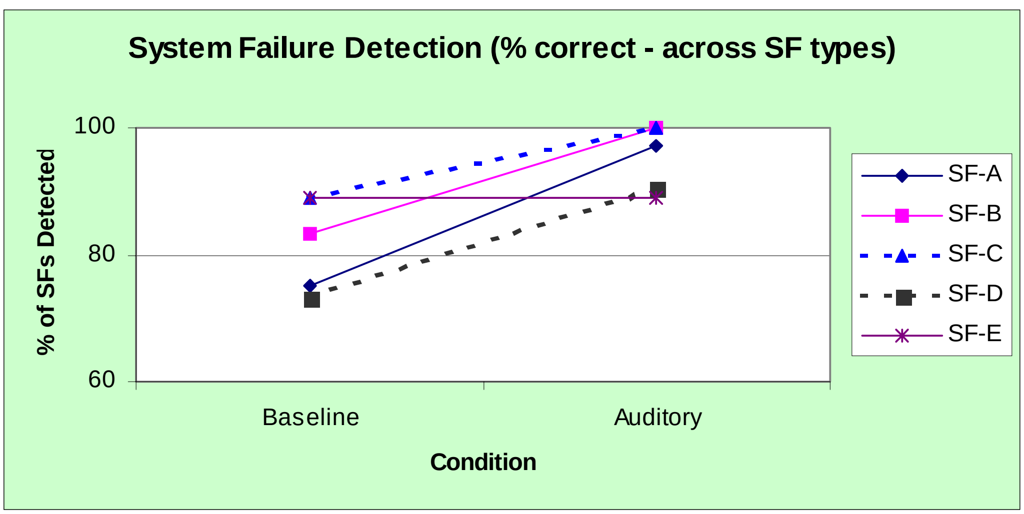

Figure R2. Overall SF detection rate for Baseline and Auditory conditions.

A planned one-tailed comparison of the condition means showed a significant overall main effect of Condition, $t(17) = 2.79$, $p < .01$, suggesting that the detection rate is generally higher under the Auditory condition than under the Baseline condition. Further one-tailed planned contrasts revealed higher detection rates for the Auditory condition than the Baseline condition for SF_A, [$t(17) = 3.06$, $p < .01$], for SF_B, [$t(17) = 2.38$, $p < .05$], for SF_C, [marginally significant: $t(17) = 1.46$, $p = .08$], but not for SF_D [non-significant trend: $t(29) = 1.29$, $p > .10$], nor for SF_E, a t-test, $t(17) = 0$, $p = 1.0$. These results are summarized in Table R1 (left columns).

Table R1. A summary of detection rate comparisons between the Baseline, Auditory, and Automation conditions across each SF Type.

Detection Rate						
	Baseline	Auditory	Baseline	Automation	Legend	
SF_A		***		***	$p < .01$	***
SF_B		**		***	$p < .05$	**
SF_C		*			$p < .10$	*
SF_D						
SF_E						

The summarized results in Table R1 reveal that the Auditory condition generally performs better than the Baseline condition in detection rates. However, as seen in SF_D and SF_E, the Auditory condition doesn't really help SF detection performance in those cases where the operator is engaged in loitering and inspecting a target (SF_D is for TOOs and SF_E is for CTs). This is possibly because these activities, in contrast to routine flight, result in a total cognitive lockout, such that the pilot is either unaware of, or intentionally chooses not to deal with, the sound announcing the system failure.

Baseline vs. automation. The data appearing in Figure R3 presents the participants' detection rate in detecting SF_A through SF_E, across the Baseline and Automation conditions.

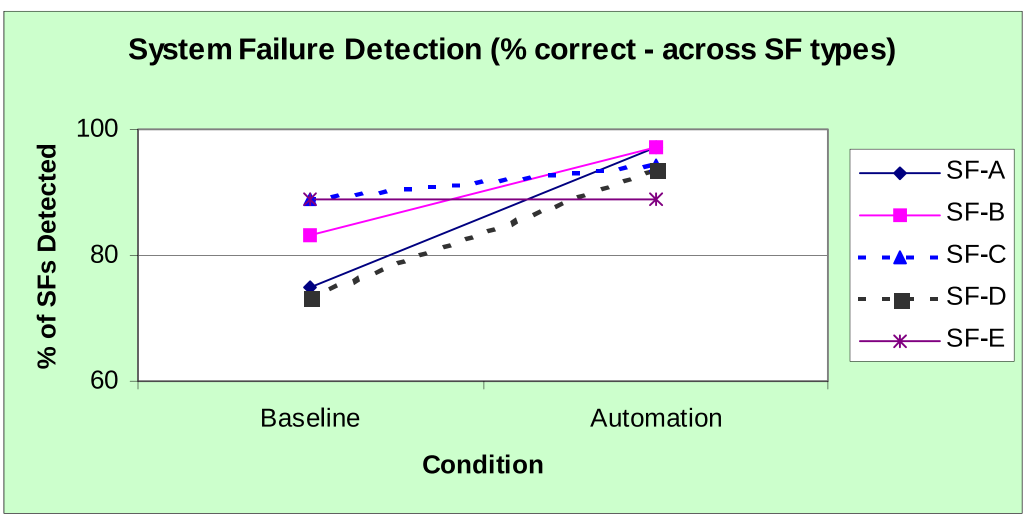

Figure R3. Overall SF detection rate for Baseline and Automation conditions.

A one-tailed planned comparison of the condition means showed a significant main effect of Condition, $t(17) = 3.09$, $p < .01$, suggesting that the detection rate is generally higher under the Automation condition than under the Baseline condition. Further planned contrasts (one-tailed for SF_A, SF_B, and SF_C since these differences were expected, and two-tailed for SF_D and SF_E since no differences were expected) revealed higher detection rates for the Automation condition than the Baseline condition for SF_A, [$t(17) = 2.68$, $p < .01$], for SF_B, [$t(17) = 1.76$, $p < .01$], but not for SF_D, [non-significant trend: $t(29) = 1.67$, $p > .10$]. Contrasts for SF_C, [$t(17) = .56$, $p = .29$] and SF_E, [$t(17) = 0$, $p = 1.0$], revealed no significant increase in detection rate from the Baseline condition to the Automation condition. These results are summarized in Table R1 (right columns).

These results reveal that the Automation condition generally supports better system monitoring performance than does the Baseline condition. However, as seen in SF_D and SF_E, the Automation condition doesn't really help performance in cases where the operator is engaged in loitering and camera zooming. As with the previous results from the Baseline-Auditory comparison, we believe this is probably due to simultaneously dealing with loitering and inspecting a target (SF_D is for TOOs and SF_E is for CTs), which results in a total cognitive lockout.

SF type. The overall differences between SF Types were analyzed when collapsed across all three conditions to see if SF_B, which occurred during periods where there were no other cognitively challenging tasks, would show higher detection rates than the other SF types, independent of the level or kind of support. However, an overall one-way analysis of variance, $F(4, 258) < 1.0$, revealed no significant differences in SF Type, suggesting that the detection rates were high (> 86%) across all five SF types.

System Failure (SF) Response Times

Baseline vs. auditory. The data appearing in Figure R4 presents the participants' response times to SF_A through SF_E, across the Baseline and Auditory conditions. The response times are measured in seconds, and express the time it takes to both detect and correct system failures (data on the former measure alone were not available).

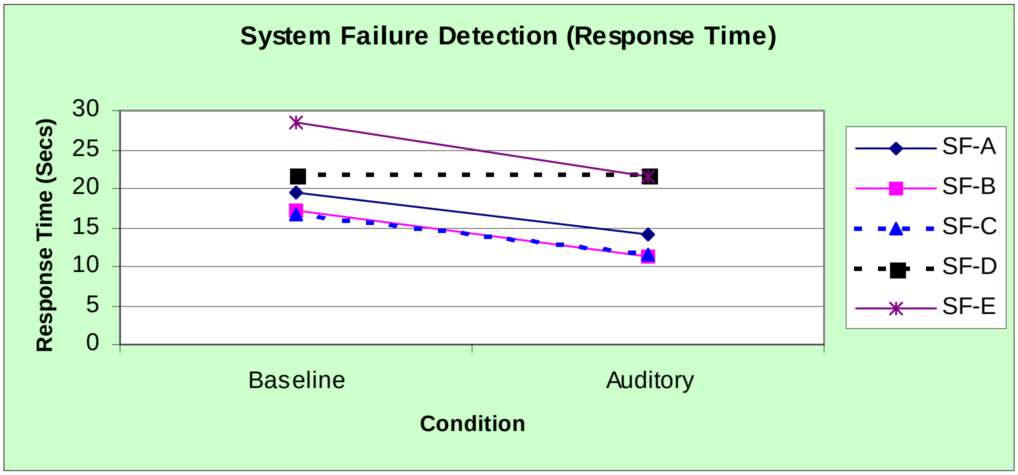

Figure R4. Overall response times to system failures (SF) between the Baseline and Auditory conditions.

A planned comparison of the condition means showed a significant overall main effect of Condition, $t(17) = 2.69$, $p = .007$, suggesting that response time is generally faster under the Auditory condition than under the Baseline condition. Further planned contrasts for each SF Type revealed faster response times for the Auditory condition then the Baseline condition for SF_A, [$t(17) = 2.02$, $p = .03$], for SF_B, [$t(17) = 3.00$, $p = .004$], and for SF_C, [$t(17) = 2.33$, $p = .016$]. On the other hand, SF_D, [$t(17) = .002$, $p = .50$], and SF_E [$t(17) = 1.00$, $p = .16$], showed no significant decrease in response time between the Baseline and Auditory conditions. These results, reinforcing the detection rate findings, are summarized in Table R2 (left columns).

Table R2. A summary of response time comparisons between the Baseline, Auditory, and Automation conditions across each SF Type.

Response Time						
	Baseline	Auditory	Baseline	Automation	Legend	
SF_A		**			$p < .01$	***
SF_B		***			$p < .05$	**
SF_C		**			$p < .10$	*
SF_D						
SF_E						

The summarized results in Table R2 reveal that the Auditory condition generally performs better in response time tasks than the Baseline condition. However, as we saw with detection rate data, this benefit is lost during SF_D and SF_E (i.e., dealing with loitering and inspecting a target simultaneously).

In addition to response times, the accuracy of the response was also recorded. This was measured by analyzing whether or not the participant correctly indicated which SF had failed, and where (ownship coordinates) that SF occurred. A two by five between-subjects analysis of variance showed no significant difference, $F(1, 149) < 1.0$, between the Baseline and Auditory conditions. Accuracy was consistently high (> 94%) across all conditions.

Baseline vs. automation. The data appearing in Figure R5 presents the participants' response times to SF_A through SF_E, across the Baseline and Automation conditions.

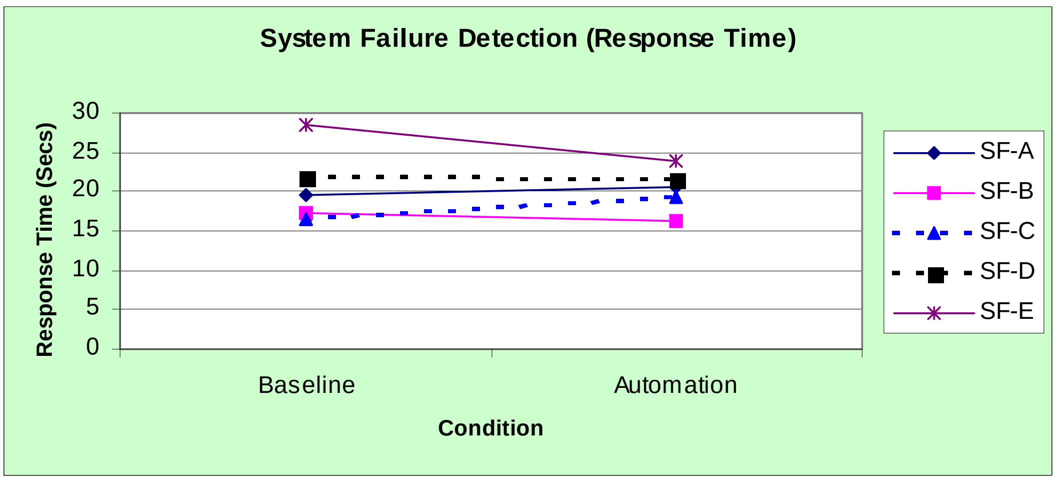

Figure R5. Overall response times to system failures (SF) between the Baseline and Automation conditions

A two by five within-subjects analysis of variance showed no significant main effect of Condition, $F(1, 17) < 1.0$ on RT, suggesting that the Automation condition did not support faster SF response times than the Baseline condition.

Regarding response accuracy, a two by five between-subjects analysis of variance showed only a non-significant trend, $F(1, 151) = 2.45$, $p = .12$, for lower accuracy in the Automation condition relative to the Baseline condition.

SF type. Figure R6 shows the overall differences between SF Types collapsed across all three conditions. We analyzed this to see if SF_B, which occurred during periods where there were no other cognitively challenging tasks, would show faster response times than the other SF types, independent of the level or kind of support.

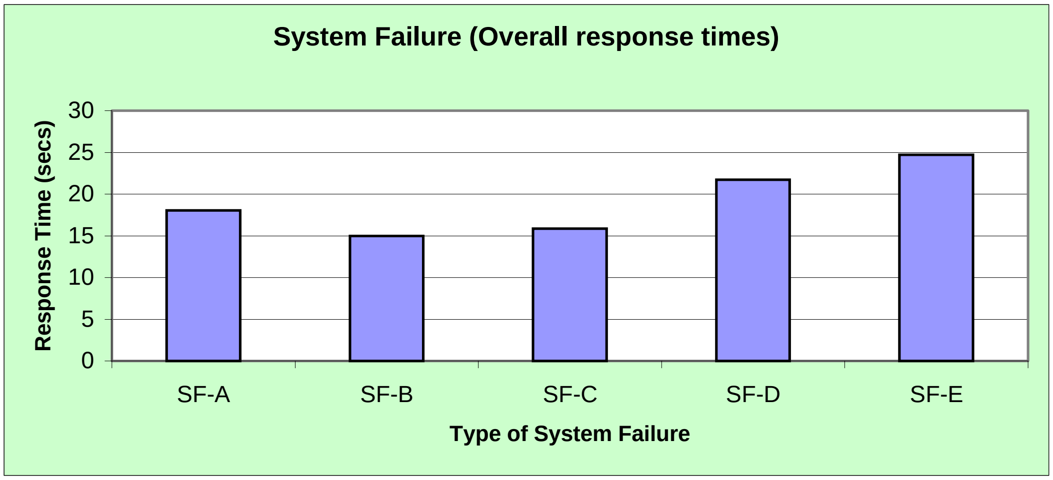

Figure R6. Overall response times to SFs collapsed across all three conditions.

A one-way analysis of variance showed a significant main effect of SF Type, $F(4, 265) = 7.73$, $p < .0001$. Using SF_B as the baseline (since it occurred alone), paired t-tests (one-tailed) for each SF Type overall mean revealed a 3-second slower response time for SF_A, [$t(53) = 2.62$, $p = .006$], a 7-second slower response time for SF_D, [$t(53) = 4.52$, $p < .0001$], and a 10-second slower response time for SF_E, [$t(53) = 3.94$, $p = .0001$]. SF_C showed no significant difference in response time, $t(53) = .95$, $p = .17$.

These results are predictable since SF_A, SF_D, and SF_E, all occur during time periods where the operator is involved in other tasks (i.e., initiating heading or analyzing targets), and these other tasks cause enough interference to disrupt the SF task. However, since SF_C occurs *before* the TOO, its response time is apparently not affected. That is, once the pilot begins to deal with the system failure he or she presumably completed it without delaying to initiate identification of the target of opportunity, or, for most TOOs, without even noticing that the TOO had appeared, an issue we address in the next section.

Searching Task

Targets of opportunity (detection rate). Figure R7 shows the percentage of TOOs correctly detected for each condition. First we were interested in seeing whether or not offloading SFs and instructions to the auditory channel had any beneficial effects on detecting TOOs. A one-tailed planned comparison, [$t(17) = 1.94$, $p < .05$], showed that offloading to the Auditory channel did offer an 8% benefit to detection rate performance for TOOs in general, although we see below that this auditory benefit is restricted to times when a SF occurred just prior to the TOO.

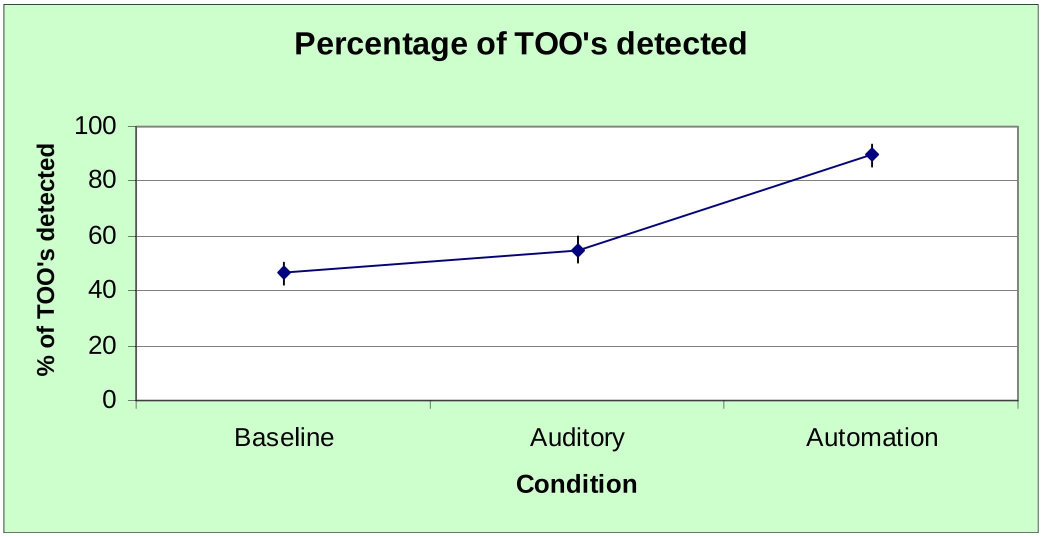

Figure R7. Percentage of TOOs correctly detected for each condition. Error bars report 95% confidence intervals.

A second analysis examined whether automating the tracking task would provide a benefit in detecting TOOs. Under the Automation condition, the UAV always flew a straight-line path to the next command target. Each of the TOOs were located along the straight-line path; therefore, in the Automation condition, every TOO came into the "window of view". On the other hand, only about 72% of the TOOs came into the "window of view" in the Baseline and Auditory conditions due to tracking error by the operator.

A planned comparison, [$t(17) = 8.84$, $p < .0001$ showed that the Automation was indeed superior to the Baseline condition in detection rate performance. As mentioned, tracking error led to many lost opportunities in detecting TOOs during the Baseline condition, so part of the reason for the poorer Baseline performance could be simply that the operator missed more opportunities. However, of the TOOs which *did* come into the "window of view," there was still a significant difference, [$t(17) = 3.57$, $p = .001$], between detection in the Baseline condition (47 out of 72, or a 66% hit rate) and the Automation condition (90 out of 100, or a 90% hit rate). Therefore, at least some of the performance improvement in the Automation condition is due to workload alleviation through automating the tracking task.

<u>Accuracy of target description</u>. With regards to accuracy in reporting target information, a one-way analysis of variance showed no significant main effect of Condition, $F(5, 74) < 1.0$. That is, the percentage of correct responses to target questions was similar across all three conditions for TOOs in general. This equivalence was also observed for TOOs that occurred with SFs (i.e., SF_D).

<u>Training</u>. A three by two within-subjects analysis of variance showed no significant differences between the means of the first five legs and the last five legs, [$F(1, 17) = 1.30$, $p = .27$], suggesting that there was no "training" effect; that is, the participants did no better during the second half of the experiment than they did during the first half.

Targets of opportunity (response time). Program uncertainty regarding the image camera orientation during flight made it impossible to assess precisely when each TOO came into view. As such, we were not able to measure the latency to detect the TOO. Instead, response time measures the time between when the loiter was entered and finishing the report. As with TOO detection results, it is important to see whether the response times for the Auditory (16.37 secs) or Automation (16.13 secs) condition might improve over the Baseline (14.51 secs) condition. A one-way within subjects analysis of variance showed no significant difference between conditions $F(2, 51) < 1$.

The fact that the Auditory condition did not shorten response times to TOOs is interesting to note, because one might expect that having an auditory alarm for SFs might aid the operator in examining the TOOs (e.g., he or she doesn't have to constantly look over to visually check the systems for failures). However, as noted in the previous SF results, analyzing a target is such a cognitive challenge that the operator appears to be locking out all other tasks while focusing on this one, independent of the modality of SF alerting or the presence of automation.

In order to further highlight this finding, we analyzed specifically just those response times to TOOs that occurred *with* SFs (recall that SF_D occurs during a TOO loiter). Breaking down the data this way would indicate if there was any Auditory benefit when the two tasks occurred at the same time. A one-way between-subjects analysis of variance showed no significant main effect for Condition, $F(2, 41) < 1$. This suggests that the response times for TOOs during SF_D was the same across all three conditions.

Therefore, when combining these results with the previous SF results, it is appropriate to conclude that having auditory alarms built into the SFs did not improve performance for either a) SF detection rate, b) SF response time, c) SF accuracy, d) TOO response time, or e) TOO accuracy, when SFs occurred concurrently with TOOs.

As mentioned, this is probably happening because the task of manipulating the image camera and analyzing targets is so difficult that timesharing becomes virtually impossible. To examine this point, we have calculated the response times for TOOs occurring alone (when there was no SF_D) and the response times to SF_B (recall that these occur when nothing else is happening other than monitoring the flight path), and compared the sum of those two **single** tasks to the total response time when TOOs and SF_D occur as a **dual** task. This latter value, which we describe as TOO / SF_D dual task, was calculated as the time from initial loiter around the TOO to the concluding response to either the TOO or SF_D, *whichever occurred last*. Thus, it is a measure of the total time to complete both tasks in the dual task context. Figure R8 shows these comparisons.

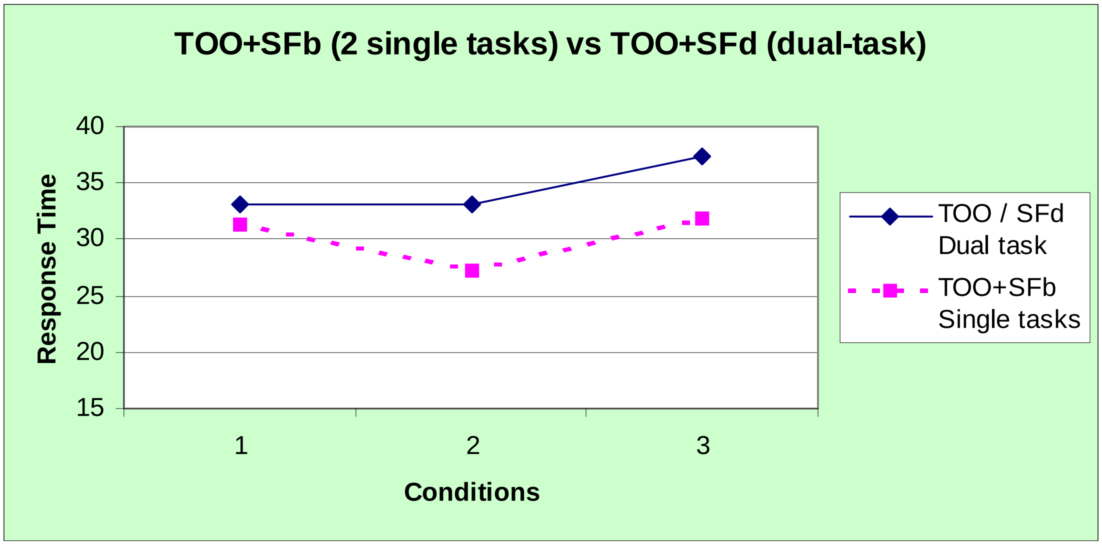

Figure R8. A comparison of response times between [TOOs that occur alone + SFs that occur alone] vs. [TOOs that occur together with SFs]. Error bars report 95% confidence intervals.

One outlier was removed from the data because it fell well beyond the third SD from the mean. Not surprisingly, it takes just as long for the operator to perform the TOO and SF_D tasks simultaneously as it does to perform them separately. In fact, in the Auditory condition, [t(28) = 1.46, p = .08], and in the Automation condition, [t(31) = 1.32, p = .097], it takes 5-6 seconds *longer* for the simultaneous tasks to be completed than if they were each performed alone. This delay probably results because the operator is switching attention back and forth between the tasks, and losing time in the scanning process. The combination of all these results clearly shows that the dual-task of analyzing targets and dealing with system failures is too difficult to allow for timesharing between the tasks.

So far, the results have suggested that timesharing is not occurring when SFs occur *during* TOOs. In order to examine the processing in the inverse situation, when a TOO appears during a SF, we examine what happens when a SF occurs just *before* a TOO (i.e., SF_C). This is done to see if having to deal with a SF will affect the accuracy of detecting TOOs. Figure R9 compares the percentage of correct TOO detections when SF_C occurs to the percentage of correct TOO detections when there is no SF_C (i.e., the majority of TOOs).

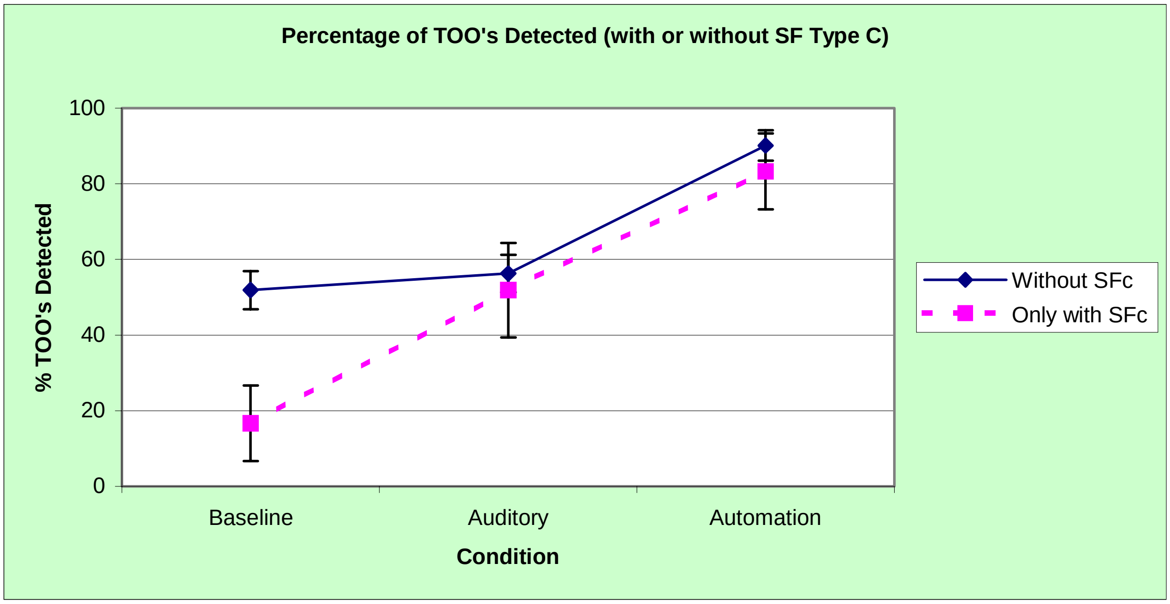

Figure R9. Percentage of TOOs correctly detected (with or without SF_C) across all conditions. Error bars report 95% confidence intervals.

As mentioned previously, it was established that the Auditory and Automation conditions improve overall TOO detection performance relative to the Baseline condition. What is interesting here is that planned comparisons show that only in the Baseline condition, [$t(17) = 3.61$, $p = .001$], is there a significant difference between detection rates of TOOs with SF_C and TOOs without a SF_C. This difference is absent in the Auditory condition, [$t(17) = .93$, $p = .18$], and in the Automation condition, [$t(17) = 1.04$, $p = .16$]. Note that the advantage the Auditory condition holds over the Baseline condition in the overall detection of TOOs is eliminated when SF_C are removed from the data [$t(17) = .89$, $p = .19$]. On the other hand, when a SF_C occurs, the Auditory condition results in a 34% improvement in detection rate over the Baseline condition.

Intuitively, one might posit that the Auditory condition should perform better than the Baseline condition when TOOs occur *without* a SF_C, because having an auditory alarm for SFs relieves some of the visual scanning workload, thus allowing the operator to focus more heavily on detecting targets. However, the results show that the Auditory condition only holds a TOO accuracy advantage over the Baseline condition *with* a SF_C. This discrepancy may be explained by examining the data from SF response times; recall in Figure R5, that under the Auditory condition, the operator tends to respond 5-6 seconds faster to the SF_C than in the Baseline condition. Therefore, the fact that the operator completes the SF task more quickly should allow him/her to return to the target search task more quickly, and thus improve the odds of catching the TOO in the "window of opportunity."

A two by three between-subjects analysis of variance showed no significant difference of response times to TOOs during a SF_C across Conditions, $F(2, 48) = 1.6$, $p = .21$.

Command targets. As with TOO response times, it is also important to see whether the Auditory or Automation conditions might improve response times to the command targets at the end of each leg. Because detection was not an issue here (the command targets were very obvious as shown in Figure M1), response time measured the time between when a loiter was triggered, until the final report was given. A one-way within subjects analysis of variance showed no significant main effect for Condition, $F(2, 51) < 1$. As mentioned in the context of the TOO response time results, analyzing a target is such a cognitive challenge that the operator appears to be locking out all other tasks while focusing on this one. This appears to eliminate any auditory advantage that might occur in a less strenuous cognitive task.

With regards to accuracy in reporting target information, a one-way analysis of variance showed no significant main effect of Condition, $F(5, 92) = 1.07$, $p = .38$. That is, the percentage of correct responses to target questions was similar across all three conditions for CTs in general, and also for CTs with SFs (i.e., SF Type E).

As with the TOO data, we also analyzed the response times to CTs that occurred *with* SFs (recall that SF_E occurs during a CT loiter). A one-way between-subjects analysis of variance showed no significant main effect for Condition, $F(2, 47) < 1$, indicating that the response times for CTs during a SF_E were the same across all three conditions. Therefore, when combining these results with the previous SF results, it is fair to say that having auditory alarms built into the SFs during CTs did not improve performance for either a) SF detection rate, b) SF response time, c) SF accuracy, d) CT response time, or e) CT accuracy.

As mentioned with regard to the TOO data, this absence of benefit probably results because the task of analyzing targets is so difficult that timesharing is impossible. Similar to the analysis portrayed in Figure R8 for TOOs, here for CTs we calculated the response times for CTs occurring alone (when there was no SF_E) and the response times to SF_B (recall that these occur when nothing else is happening). The sum of those two *single* tasks were compared to the total response time to complete both tasks when CTs and SF_E occur as a *dual task* (CT / SF_E dual task). Figure R10 shows these comparisons.

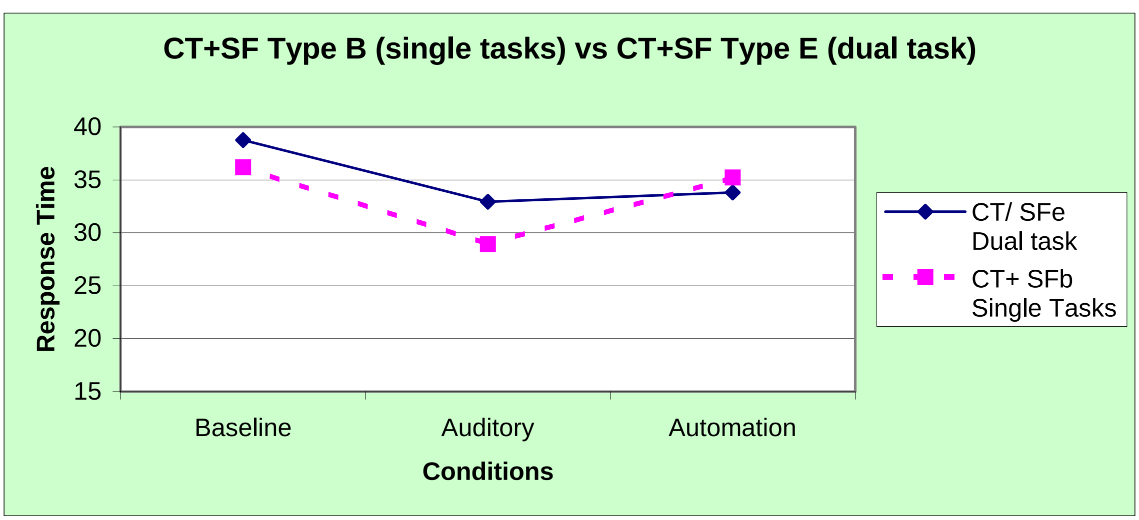

Figure R10. A comparison of response times between [CTs that occur alone + SFs that occur alone] vs. [CTs that occur together with SFs].

Again, it takes just as long for the operator to perform the CT and SF tasks simultaneously as it does to perform them separately. The combination of all these results clearly shows that the dual-task of analyzing targets and dealing with system failures is too difficult to allow for timesharing between the tasks, and given this difficulty, CT performance does not avail itself of the resources released in the Auditory condition.

Repeats. The participants were required to read, or listen to, command target instructions at the beginning of each leg, describing both the coordinates and the needed report. These instructions were only available for 10-15 seconds, and could be repeated if necessary by a keypress request. Figure R11 shows how often these instructions were repeated across the three conditions. These data are examined to see if memory plays a role in multi-tasking.

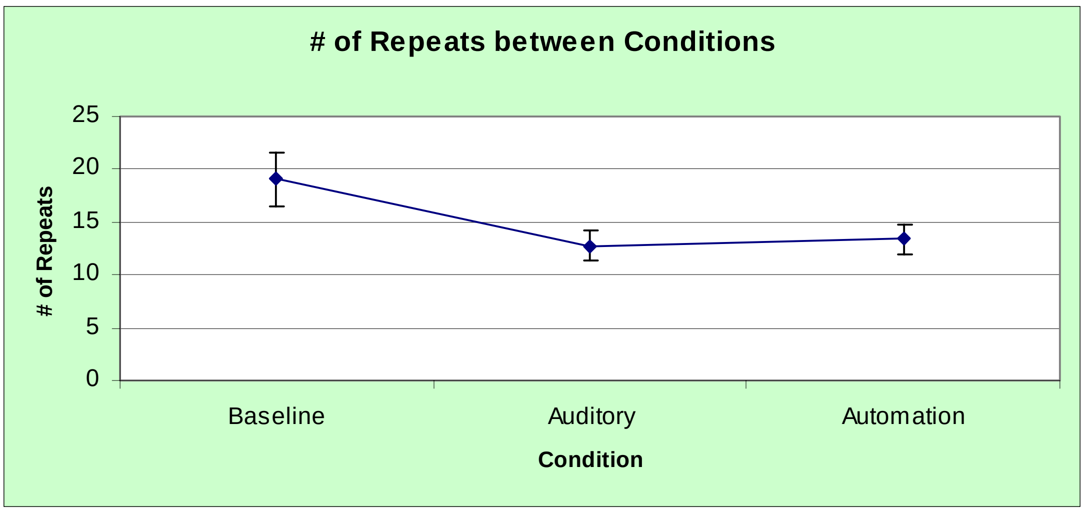

Figure R11. Number of times participants repeated the instructions across all conditions.

A Friedman analysis of variance shows a significant difference between the three conditions, [$\chi^2_F = 7.19$, $p < .05$)], suggesting that the Baseline condition generates more repeats than the other two conditions.

The reason for the fewer repeats in the Auditory condition can be explained based on a multiple resource interpretation. In the Auditory condition, the operator can effectively process the instructions during the 15-seconds that are available, because they are using separate resources (i.e., listening to instructions while visually monitoring). This causes less interference than in the Baseline condition, where the operator must use visual resources for both tasks. In the Baseline condition, the operator *must* divide his/her visual attention between reading instructions and visual monitoring (SFs, 3D view, and 2D map). To be performed effectively, all of these tasks require foveal vision, and the visual angles (7 - 23.5 degrees) between the tasks prevents parallel processing. This increase in visual competition is presumably compensated for by more repeats, so that navigation and CT reporting performance did not suffer.

Fewer repeats in the Automation condition is assumed to be due to the fact that the operator does not have to manually track, and thus does not have to recheck the coordinates for the next command target.

Subsequent analysis focused on exactly *when* the operator is pressing the Repeat button along each leg. Bear in mind that each instruction contains two qualitatively different pieces of information: 1) the location of the CT, which needs to be established at the beginning of each leg, and 2) the nature of the report once the CT is reached, which is located at the end of each leg. Figure R12 shows the timeline of Repeats through an average mission "leg" (10 legs per mission) across the three conditions.

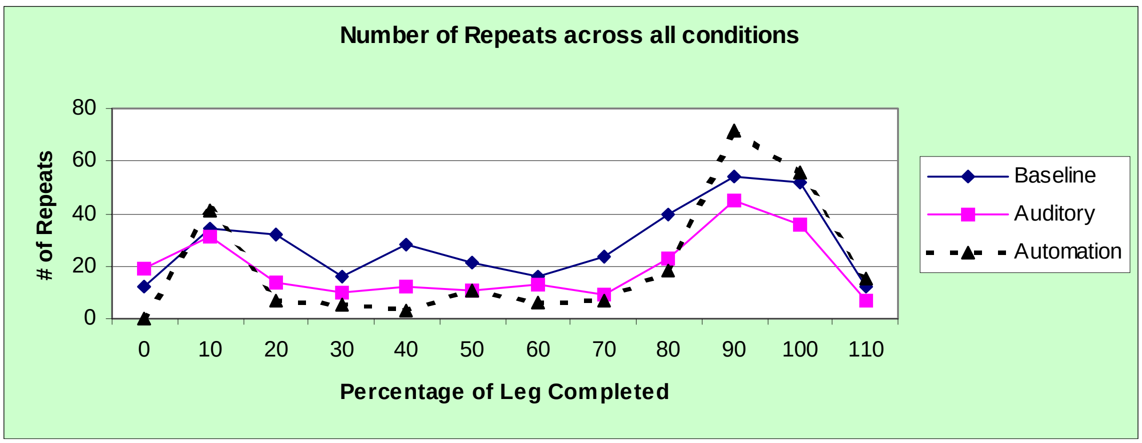

Figure R12. The number of repeats averaged over 10 mission "legs" in the Baseline condition, across the three conditions. Values along the x-axis over 100% indicate that the operator overshot the CT.

All three conditions show the general increase in repeats as the CT is neared, reflecting the need to refresh memory of the instructions just prior to CT inspection. Furthermore, a Friedman analysis of variance shows a significant difference between the three conditions, [$\chi^2_F = 6.12$, $p < .05$)], suggesting that the distributions of the three conditions are not the same.

Two points of interest in Figure R12 reflect these differences. First, under the Automation condition, the pilots appear not to check the instructions very often in the middle of the mission leg. However, towards the end, they tend to check them more often. This makes sense because in the Automation condition, the operator doesn't need to recheck the coordinates after a SF or TOO, since they know that the UAV will go there without any control manipulation. However, they probably check more towards the end of the leg than in the Baseline condition because they haven't refreshed their memory of the required report nearly as often during the leg as in the Baseline and auditory conditions.

Second, the Auditory condition appears to follow a similar pattern as the Baseline condition, albeit with fewer overall Repeats as noted in the context of Figure R11. This equivalence in time distribution is predictable, since both these conditions require manually flying the UAV, and thus the operator needs to recheck the coordinates after dealing with a SF or a TOO.

Discussion

This simulation was designed to provide data on pilot performance measures of workload when controlling a single RPV, within which was embedded a set of partially overlapping tasks. At the highest level of goal structure, three tasks were presented:

1. Navigate to a command target and report its features
2. Monitor for targets of opportunity (TOO) while en route, and report these if spotted.
3. Monitor for on-board system failures.

Each of these were in turn associated with one or more channels of perceptual information and varying levels of cognitive and response demands. Furthermore, certain of these tasks could be broken down into subtasks as shown in Table D1, reproduced below.

Table D1. Task analysis.

1. Navigation:
    - 1.1 Read (or hear) CT location
    - 1.2 Establish coordinates (by orienting vehicle by joystick control or typing)
    - 1.3 Monitor heading toward CT location on 2D map (and re-orient if necessary)
    - 1.4 Refresh memory for location and final report
    - 1.5 Inspect image
        - 1.5.1 Enter loiter
        - 1.5.2 Zoom in
        - 1.5.3 Adjust camera orientation
        - 1.5.4 Count identify and/or assess cardinal orientations
        - 1.5.5 Verbal report of content.

2. TOO task:
    - 2.1 Monitor 3D display
    - 2.2 Inspect image if target located
      (see 1.5 for subtasks)

3. System Failure
    - 3.1 Monitor for System failures
    - 3.2 Identify failure
    - 3.3 Keyboard data entry

While all three tasks were, to some extent, to be performed concurrently, not all phases of each task overlapped with phases of each other task, and generally, the separate phases within a task did not overlap.

In the following discussion, we first describe the evidence for interference between these tasks within the baseline condition; then we consider how this interference was modified by offloading certain task components (1.1, 1.4 and 3.1) to auditory displays, and by automating navigation (1.2 becomes keyboard data entry, 1.3 is eliminated).

Baseline Condition

In the baseline condition, clear evidence was provided that the combined demands of visual monitoring for the three tasks across three display areas (map, 3D image, system gauges), produced substantial interference. Given the relatively high degree of spatial separation across these displays (see Figure M2), along with the high acuity demands associated with all tasks, this interference is to be expected (e.g., Liao & Moray, 1993). For both the TOO and the SF monitoring tasks, detection accuracy was degraded below the 100% value that would be expected under focused attention conditions, even during the phase when both tasks required only monitoring (SF b: see Figure R3 and R7).

For system failure monitoring, this interference can be attributed to the joint demands of monitoring the 3D display for TOOs and monitoring the map to assure that the appropriate heading was selected toward the target. For TOO monitoring, interference can again be attributed to map heading monitoring, as well to as the demands of monitoring for system failures. Furthermore, an additional source of TOO detection rate decrement in the baseline condition resulted because tracking deviations from the straight line path between command targets prevented some of the targets (28%) from ever appearing in the 3D window. But a substantial portion of the remaining 72% that did appear were still overlooked, since only 45% of all targets were detected. That is, the miss rate of visible targets was 45/72=62%.

Such monitoring-monitoring interference effects could be predicted by all three models of time sharing: single channel, single resource and multiple resource. Clearly all tasks compete for the same visual resources, a state consistent with single and multiple resource models as well as those predicting single channel processing (here “single channel” in access to foveal processing).

The baseline data also provide clear evidence for a quantitative increase in interference attributable to the cognitive/motor demands of image inspection and interpretation, above and beyond the pure perceptual demand of monitoring. This evidence is provided from two sources: (1) the loss in SF detection performance (both detection rate and speed) from SFb (occurring during pure monitoring) to SFd (occurring during ongoing 3D image inspection; see left side of Figure R3, R4). (2) The decrease in TOOs detected when those TOOs arrive just after a system failure has occurred (SFc: see left side of Figure R9). Thus some aspects of the added visual cognitive and motor components of both tasks appear to disrupt the visual monitoring of the other. This disruption is not just in the form of a minor delay, but rather is a complete absence of noticing. That is, attention appears to be diverted for a long enough time that the visual evidence of the event (SF or TOO) is gone, by the time attention is returned, thereby reflecting a sort of cognitive tunneling or attentional fixation (Fadden, Ververs, & Wickens, 2001). Our analysis of the auditory condition below helps establish that it may be the cognitive components that are responsible. This cognitive-related decrement is substantial, leading to a reduction in SF detection rate from 85% to 72%, and a reduction in detection rate of the more difficult TOO detection from 50% to 17%. A final analysis reported in Figure R8 provides some further evidence for “single channel” behavior, when a TOO was detected prior to an SF occurrence. Here the combined time to complete both tasks was actually longer than the sum of the estimated times to complete each one alone. Had pilots been performing some aspects of both tasks in parallel, this combined measure would be predicted to be less.

The system failure data also provided clear evidence that the strong perceptual and cognitive demands associated with the initial phase of selecting the heading and initiating the trajectory (when Sf-a occurred) as well as the final command target phase (when Sf-e occurred) substantially disrupted SF evaluation performance, although whether the longer RTs in the period are a consequence of slower detection or longer keyboard entry (of failure type and current coordinates) cannot be discriminated from the current data. TOO’s were not present during these periods, so their detection data did not offer any relevant performance metrics.

It is impossible from the current data to assess how much the navigational task itself (task 1.3) was disrupted by monitoring for the two kinds of events, since no independent “single task” version of tracking was recorded. However, we do note that navigation was far from perfect, compared to its performance level in the automated condition, as seen below, and as reflected in the number of TOOs that failed to appear in the 3D display window. Furthermore, a second part of the navigation task, assessing (or remembering) the coordinates (task 1.4) appeared to be less than perfect. Repetitions of this information were requested early in each mission leg (Figure R12), at a time when the nature of the information needed was more likely to be the coordinates than the final report requirements (although this could not be distinguished with certainty). It is impossible to establish, however, the extent to which these repeat requests were the result of interference from other visual tasks, or simply the result of working memory failures.

Auditory Offloading

Against the backdrop of substantial visual interference between monitoring and cognitive/control aspects of the various tasks, experimental interest was directed to establishing the extent to which task interference could be reduced by re-distributing some aspects of the tasks to the auditory channel. To the extent that multiple resource theory accurately captures human multi-task performance, and that task demands are visual (rather than cognitive or motor), greater performance benefits should be observed. In particular, by announcing system failures with an auditory signal, this should relieve pilots of the need for continuous visual monitoring (task 3.1), freeing those visual resources for better performance on navigation (task 1) and target monitoring (task 2). Auditory presentation of coordinates and CT instructions were expected to have corresponding effects.

In fact, the most noticeable benefits of auditory offloading were observed on the tasks that were directly served by the auditory channel, rather than on concurrent tasks of navigating and target monitoring. In particular, SF monitoring was improved in both speed and accuracy, by its auditory annunciation (Figures R2, R4, Tables R1 and R2), so long as the pilot was not engaged in detailed image inspection. That is, pilots could listen for the occurrence of system failures, while they were looking for TOOs, or looking at the map to establish their trajectory (parallel cross-modal monitoring). In contrast, the auditory benefit was not observed in SFd and Sfe. The benefit to detecting system failures during monitoring phases (Sfa, b, and c) is clearly consistent with multiple resource predictions. However the reduction, or absence, of an auditory benefit during these image inspection phases suggests that the high **cognitive** demands of those tasks (1.5.2-1.5.5 in Table D1) may be such as to negate most advantages of visual offloading. That is, it is the cognitive, not the perceptual demand of the target task that is the source of most interference here, and this cognitive competition is independent of input modality.

In contrast to SF monitoring, the data provide little evidence that TOO monitoring was directly supported by auditory display of system failures. While this auditory offloading did provide a very modest and statistically significant benefit to TOO detection rate (Figure R7), further analysis suggested that this benefit was entirely an indirect one, related to the shorter latency of completing the SF task (compared to the baseline) when it was delivered auditorially just prior to a TOO appearance (i.e., SFc; see Figure R9). The fact that pilots noticed the SF sooner because of its auditory delivery, allowed them to complete the response sooner, and return more quickly to more focal monitoring of the 3D image display where the TOOs were located. As a consequence, fewer of those TOOs passed unnoticed.

Finally, although there was no evidence that auditory delivery of command target instructions (task 1.1) assisted navigation directly, it did apparently provide a more effective way of delivering navigational information in this highly visual environment. Such evidence was provided by the fewer number of repeats requested. It may be that the auditory delivery of linguistic information made it more enduring because of higher compatibility (Wickens, Sandry, & Vidulich, 1983). Alternatively, it may be that visual delivery forced pilots to share visual attention between the message box and other displays, hence leading to a less effective retention of the visually displayed instructions, and requiring more repeats. The latter interpretation again is directly based upon multiple resource benefits.

Further exploration of possible auditory benefits were provided by the single channel analysis shown in Figure R8. Here the data actually revealed a greater tendency toward single channel behavior (i.e., a greater amount by which the combined RT exceeded the sum of the separate RTs) in the auditory than in the baseline condition. Such a finding reinforces the evidence, provided by the lack of auditory benefits during SF-d and Sf-e, that auditory offloading has little benefit when the cognitive demands of one or both tasks are extremely high. That is, the high competition for cognitive resources dominates any benefits for separate peripheral resources.

Automation Benefits

While the goal of auditory offloading was to **distribute** the same task demands across different resources, the goal of automation implementation was to **reduce** some of those demands associated directly with the navigation task 1.3 by eliminating the pilot's need to monitor the RPV course, once the initial coordinates had been typed in. ("Elimination" assumes perfect automation, an issue whose implications were not examined in the current study, but see Wickens, 2000 for discussion of its importance). Automation was found to exert three different classes of effects, only the third of which was of direct experimental and practical interest from a pilot performance perspective. (1) Since the automation was programmed here to be perfect, it substantially reduced tracking error (to essentially zero). (2) Because of (1), the simulated UAV precisely followed the simulation trajectories, and therefore brought every TOO within the viewing window, availing a greater number of those TOOs to be detected relative to the baseline and auditory conditions. This effect was in some respects an experimental artifact, since in real environments, TOOs would not be expected to exclusively lie under the flight path. (3) Automation produced some real benefits to concurrent task performance, the primary interest of our analysis. We discuss these as follows.

It became apparent from the comparison between the baseline and automation condition performance on both detection tasks that flight control (task 1.3) imposed substantial resource demands. Its removal as a task by automating improved SF monitoring detection rate, both during the initial trajectory orientation phase (Sf-a) and during the pure monitoring phase (SFb), although there were no benefits in detection speed.). The detection rate benefit was very slight -- a non-significant trend -- for SF-c, which occurred at a time well into the flight leg, at which the manually oriented trajectory in the baseline condition was probably well aligned and required little further adjustment. Benefits were neither predicted nor found for those system failures (SFd and Sfe) that occurred during loiter, where the level of automation did not differ from the baseline (since both conditions provided automated loiter flight).

Automation also had a substantial influence on improving TOO detection from 42% in the baseline condition to 90% when flight trajectory was automated. It is recalled that part of this benefit is simply related to the artifact of more precise autopilot tracking of the legs, bringing more targets into the display window. However even adjusting for this difference, the baseline condition still supported a detection rate of observable targets of only 66% compared to 90% for automation. Thus the visual resources relieved from path monitoring, and occasional adjustment, were applied productively in monitoring the 3D camera image for the low salience targets.

Finally, it appears that automation provided some relief from rechecking the instructions, as shown in Figures R11 and R12. Such a relief, as with the improved navigational accuracy, is a direct reflection of the automation functionality. Once the target coordinates were entered at the outset, there was little need to check the heading, one purpose of the instruction refresh. Only toward the end of the flight, just prior to reaching the command target was it necessary to refresh memory as to what specifically was to be reported there.

Implications for Models of Divided Attention

The current data do not serve to “prove” or “disprove” any of the three models of multiple task performance or divided attention described at the outset: single channel theory, single resource theory and multiple resource theory. Some aspects of the data can be accounted for by all three, whereas there are other aspects that suggest certain models may apply in some circumstances, but not in others. As we have noted, some evidence was supplied for the benefits of auditory offloading in availing more visual resources to the remaining tasks, or, in the case of the SF task, reducing the time that a system failure remained unnoticed In this case, the data suggest that the auditory offloading allowed SF detection to occur concurrently with TOO and flight path monitoring. Correspondingly, the data suggest that auditory instructional delivery allowed instructions to be processed in parallel with other visual task demands; enabling the instructions to be better encoded, and hence, less frequently needed for refreshing. Both effects support the viability of multiple resource explanations associated with perceptual channels.

The absence of auditory benefits during the very high demand period of target inspection, one associated with continuous mental rotation and spatial problem solving (Gugerty & Brooks, 2001), suggests that the dispersal of task processing across perceptual resources provides less (or even no) benefit when cognitive demands are quite high, and are assumed to be the dominate force in dictating task interference. Such a view is actually consistent with current multiple resource models that assume perceptual and cognitive activities to compete for common resources (Wickens, 2002), although these effects could equally well be explained by single resource or by single channel theory. Further examinations of the benefits of multiple resource modeling will await comparisons of dual UAV supervision between auditory and visual interfaces (Wickens, Dixon, & Chang, in preparation).

The distinctions between resource theory (whether single or multiple) and single channel theory are somewhat more difficult to draw from the current data, in part because the concept of "single channel" has so many different manifestations, depending upon which aspect of information processing is assumed to be carried out on only one task at a time (Hendy et al., 1997; Liao & Moray, 1993; Pashler, 1998; Welford, 1967). Also, single channel theory can make various predictions, depending upon the speed and timing of switching between tasks. There is little doubt that some aspects of the current simulation forced "single channel access" to foveal vision. Since foveal vision is necessary to identify the small resolution TOOs (task 2.1), as well as to process the heading information on ownship (task 1.3), and read the text in a refreshed display (tasks 1.1, 1.4), it is likely that no concurrent perceptual operations are possible. More feasibly, the pilots could be assumed to be engaged in concurrent processing of the SF displays (in peripheral vision) while the 3d image display is searched. However it is difficult to determine precisely the extent to which this is true.

The problem is that any single channel perceptual model can approximate the performance of a parallel processing model if attention switching (here characterized by visual scanning) is allowed to occur very rapidly (Townsend, 1974), and since visual scanning was not measured in the current data, these distinctions are difficult to make. On the other hand, an extreme view of single channel processing, characterized by attentional tunneling, would predict total abandonment of one task, till another is completed. The strongest predictions of this view, in the current model, are that system failures occurring after a TOO has been detected, and particularly, after the loiter had been entered, and image inspection initiated, would not be detected until the TOO task has been entirely completed. Indeed, there is nothing in the current data that would discount this possibility, since the time that system failures remained visible clearly exceeded the time at which the TOO task was completed. Future analysis will be required to establish the extent to which total single channel behavior described the circumstances when a report was required for both monitoring tasks, or rather, some degree of more rapid task switching was invoked.

In conclusion, the current research has developed a relatively valid simulation for RPV control, and provided data which can be used to assess the interference between concurrently performed task components. In two future reports we will (a) describe the modeling of such data by computational models of task interference, (b) report the results of a dual RPV experiment.

Acknowledgments

The authors would like to express their extreme gratitude to Ron Carbonari and Jonathan Sivier, whose efforts were responsible for developing the current simulation. Dr. Mike Barnes, and personnel at Ft Huaccuca, Arizona helped support the cognitive task analysis used to develop that tasks used in the simulation.

References

Carbonnell, J. F., Ward, J. L., & Senders, J. W. (1968). A queuing model of visual sampling: Experimental validation. IEEE Transactions on Man Machine Systems, MMS-9, 82-87.

Chou, C., Madhavan, D., & Funk, K. (1996). Studies of cockpit task management errors. The International Journal of Aviation Psychology, 6(4), 307-320.

Fadden, S., Ververs, P. M., & Wickens, C. D. (2001). Pathway HUDS: Are they viable? Human Factors, 43(2), 173-193.

Gugerty, L, & Brooks, J. (2001). Seeing where you are heading: Integrating environmental and egocentric reference frames in cardinal directions judgments. Journal of Experiment Psychology: Applied, 7(3), 251-266.

Hendy, K. C., Liao, J., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load. Human Factors, 39, 30-47.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice Hall.

Keele, S. W. (1973). Attention and human performance. Pacific Palisades, CA: Goodyear Publishing Company.

Keppel, G. (1982). Experimental design and analysis. San Francisco, CA: Freeman.

Kerstholt, J. H., Passenier, P. O., Houttuin, K. & Schuffel, H. (1996). The effect of a priori probability and complexity on decision making in a supervisory control task. Human Factors, 38(1), 65-78.

Laughery, K. R. (1989). Micro SAINT – A tool for modeling human performance in systems. In G. R. McMillan, D. Beevis, E. Salas, M. H. Strub, R. Sutton, & L. Van Breda (Eds.), Applications of human performance models to system design (Defense Research Series Vol. 2, pp. 219-230). New York: Plenum Press.

Laughery, K. R., & Corker, K. (1997). Computer modeling and simulation. In G. Salvendy (Ed.), Handbook of human factors and ergonomics (2nd ed.) (pp. 1375-1408). New York: Wiley.

Liao, J. & Moray, N. (1993). A simulation study of human performance deterioration and mental workload. Le Travail Humain, 56(4), 321-344.

Meyer D.E., & Kieras, D.E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. Psychological Review, 104, 749-791.

Moray, N. (1967). Where is capacity limited? A survey and a model. Acta Psychologica, 27, 84-92.

Moray, N. (1986). Monitoring behavior and supervisory control. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), Handbook of perception and performance (vol. 2, pp. 40-1-40-51). New York: Wiley.

Moray, N. & Rotenberg, I. (1989). Fault management in process control: eye movements and action. Ergonomics, 32(11), 1319-1342.

Navon, D. & Gopher, D. (1979). On the economy of the human-processing system. Psychological Review, 86(3), 214-255.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. Cognitive Psychology, 7, 44-64.

North, R. A., & Riley, V. A. (1989). W/INDEX: A predictive model of operator workload. In G. R. McMillan, D. Beevis, E. Salas, M. H. Strub, R. Sutton, & L. Van Breda (Eds.), Applications of human performance models to system design. Defense Research Series, Volume 2 (pp. 81-90). New York: Plenum Press.

Pashler, H.E. (1998). The psychology of attention. Cambridge, MA: MIT Press.

Pattipati, K. R., & Kleinman, D. L. (1991). A review of the engineering models of information-processing and decision-making in multi-task supervisory control. In D. L. Damos (Ed.), Multi-task performance (pp. 35-68). Washington, DC: Taylor and Francis.

Raby, M., & Wickens, C. D. (1994). Strategic workload management and decision biases in aviation. International Journal of Aviation Psychology, 4(3), 211-240.

Sarno, K. J., & Wickens, C. D. (1995). The role of multiple resources in predicting time-sharing efficiency. International Journal of Aviation Psychology, 5(1), 107-130.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing II: Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition (pp. 133-185). Hillsdale, NJ: Lawrence Erlbaum.

Welford, A. T. (1967). Single channel operation in the brain. Acta Psychologica, 27, 5-21.

Wickens, C. D. (1980). The structure of attentional resources. In R. S. Nickerson (Ed.), Attention and performance VIII (pp. 239-257). Hillsdale, NJ: Erlbaum.

Wickens, C. D. (2000). Imperfect and unreliable automation and its implications for attention allocation, information access and situation awareness (Final Technical Report ARL-00-10/NASA-00-2). Savoy, IL: University of Illinois, Aviation Research Lab.

Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical Issues in Ergonomic Science, x(x) 1-19.

Wickens, C. D., Dixon, S., & Chang, D. (in preparation). Supervision of two RPV's.

Wickens, C. D., Goh, J., Helleberg, J., & Talleur, D. A. (2002). Modality differences in advanced cockpit displays: Comparing auditory and vision for navigational communications and traffic awareness (ARL-02-8/NASA-02-6). Savoy, IL: University of Illinois, Aviation Research Lab.

Wickens, C. D., Helleberg, J., Goh, J., Xu, X., & Horrey, B. (2001). Pilot task management: testing an attentional expected value model of visual scanning (ARL-01-14/NASA-01-7). Savoy, IL: University of Illinois, Aviation Research Lab.

Wickens, C. D. & Hollands, J. (2000). Engineering psychology and human performance (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D., Sandry, D., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, output, and central processing. Human Factors, 25, 227-248.